2017

Molecular Analysis Of Amyotrophic Lateral Sclerosis And Frontotemporal Degeneration Brain Tissue Identifies Disease Mechanisms Associated With Repetitive Dna Elements

Elaine Liu
University of Pennsylvania, eliu107@gmail.com

Molecular Analysis Of Amyotrophic Lateral Sclerosis And Frontotemporal Degeneration Brain Tissue Identifies Disease Mechanisms Associated With Repetitive Dna Elements

Abstract
Aging-related neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) are two fatal progressive neurodegenerative diseases that carry genetic and pathologic overlap: a hexanucleotide G4C2 repeat expansion in C9orf72 and loss of a nuclear RNA binding protein, TAR DNA binding protein-43 (TDP-43), into cytoplasmic aggregates. The C9orf72 expansion is the most common genetic cause of ALS/FTD and is associated with reduced C9orf72 expression and accumulation of toxic RNA and protein aggregates. In Chapter 2 of my thesis, using molecular analyses from human post-mortem ALS/FTD brain, I show that the C9orf72 promoter is hypermethylated within a subset of expansion carriers. C9orf72 promoter hypermethylation is associated with reduced C9orf72 pathology and may be protective in these patients. Another key feature in ALS/FTD is the characteristic pathology of nuclear TDP-43 loss in degenerating neurons. TDP-43 is a ubiquitous nuclear RNA binding protein and is heavily involved in RNA processing. TDP-43 has been shown to bind genic elements and repetitive transposable elements such as long interspersed nuclear elements (LINE). Considering that TDP-43 is a ubiquitous RNA binding protein, I hypothesize that nuclear TDP-43 loss can lead to large transcriptomic changes and may contribute to alterations in LINE elements. For Chapters 3 and 4 of my thesis, I use a novel method of subcellular fractionation and fluorescent activated cell sorting (FACS) from post-mortem ALS/FTD human brain to perform high-throughput sequencing analyses to study neuronal molecular changes. In Chapter 3 of my thesis, I use FACS coupled with RNA-seq on neuronal nuclei with and without TDP-43 to show that loss of nuclear TDP-43 is associated with large transcriptome changes and increased LINE accessibility. Furthermore, loss of nuclear TDP-43 leads to increased retrotransposition. I also extend this subcellular fractionation-FACS method to study the effects of the C9orf72 expansion in neuronal nuclei. In Chapter 4, I demonstrate that the C9orf72 expansion is linked to mild gene expression changes that reflect C9orf72 protein loss and not gain of toxic C9orf72 RNA. Through my work, I have shown disease mechanisms linked to repetitive DNA elements, in that I propose (1) the C9orf72 repeat expansion may contribute to disease primarily via a gain of toxic C9orf72 pathology (2) loss of neuronal nuclear TDP-43 may be associated with increase retrotransposon activity which may contribute to disease. Overall, my work has broadened the field of neurodegeneration in my implementation of cell-type specific molecular analyses on post-mortem brain of ALS/FTD patients to identify disease mechanisms with the intent of discovering new therapeutics and biomarkers that can be extended into the clinic.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Neuroscience

MOLECULAR ANALYSIS OF AMYOTROPHIC LATERAL SCLEROSIS AND FRONTOTEMPORAL DEGENERATION BRAIN TISSUE IDENTIFIES DISEASE MECHANISMS ASSOCIATED WITH REPETITIVE DNA ELEMENTS

Elaine Liu

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Edward B. Lee, MD, PhD
Assistant Professor of Pathology and Laboratory Medicine

Graduate Group Chairperson

_____

Joshua I. Gold, PhD
Professor of Neuroscience

Dissertation Committee:
Brian D. Gregory, PhD, Associate Professor of Biology
Kristen W. Lynch, PhD, Professor of Biochemistry and Biophysics
James Shorter, PhD, Associate Professor of Biochemistry and Biophysics
Zhaolan (Joe) Zhou, PhD, Associate Professor of Genetics

I dedicate this thesis to my parents and grandparents:

To my parents, Dawen Liu and Gloria Lig-King Y Liu. Thank you for your constant support in my long educational journey and for always pushing me to reach my full potential.

To my grandparents, Dr. Qin Liu and Yunnong Li: Thank you for fostering my strong curiosity of science and medicine at such a young age. Without both of your support, I do not think I could have made it here today.

ACKNOWLEDGMENT

Firstly, I would like to express my sincere gratitude to my advisor Dr. Edward B Lee for the continuous support of my Ph.D, for his patience, motivation, and plethora of knowledge. His guidance helped me in all my time in his lab performing research and writing of this thesis. I must thank him for taking me on as his first graduate student. Besides my advisor, I would like to thank the rest of my thesis committee for their insightful comments and encouragement.

I thank my fellow labmates both past and present for the stimulating discussions, for all the weekends we have spent in the lab, and for all the fun we have had in the past six years. I thank the University of Pennsylvania Flow Cytometry Core for good discussions when I was developing my research method. I thank my friends in the Neuroscience Graduate Group at University of Pennsylvania. I am particularly grateful to Dr. Elizabeth A Heller who gave me a chance to really learn about research when I was a mere high school student pretending to be a scientist.

I thank my fiancé, Jonathan Wong, for keeping me laughing and light-hearted through this long process, driving me to and from lab as needed and even during weekend nights. These were all integral for me making it through this graduate school process (mostly the laughter) and I could not have made it through without your support. I would also like to thank my family: my parents and grandparents, brother (Eric Liu), sister-in-law (Dr. Yi Zhang), my future mother-in-law (Betty Wong) and sister-in-law (Tiffany Wong) for supporting me throughout writing this thesis and my life in general.

Most importantly, I must thank all the patients and families who make all of this work even possible. Thank you so much.

ABSTRACT

MOLECULAR ANALYSIS OF AMYOTROPHIC LATERAL SCLEROSIS AND FRONTOTEMPORAL DEGENERATION BRAIN TISSUE IDENTIFIES DISEASE MECHANISMS ASSOCIATED WITH REPETITIVE DNA ELEMENTS

Elaine Liu

Edward B Lee, MD, PhD

Aging-related neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) are two fatal progressive neurodegenerative diseases that carry genetic and pathologic overlap: a hexanucleotide G_4C_2 repeat expansion in *C9orf72* and loss of a nuclear RNA binding protein, TAR DNA binding protein-43 (TDP-43), into cytoplasmic aggregates. The *C9orf72* expansion is the most common genetic cause of ALS/FTD and is associated with reduced *C9orf72* expression and accumulation of toxic RNA and protein aggregates. In Chapter 2 of my thesis, using molecular analyses from human post-mortem ALS/FTD brain, I show that the *C9orf72* promoter is hypermethylated within a subset of expansion carriers. *C9orf72* promoter hypermethylation is associated with reduced C9orf72 pathology and may be protective in these patients. Another key feature in ALS/FTD is the characteristic pathology of nuclear TDP-43 loss in degenerating neurons. TDP-43 is a ubiquitous nuclear RNA binding protein and is heavily involved in RNA processing. TDP-43 has been shown to bind genic elements and repetitive transposable elements such as long interspersed nuclear elements (LINE). Considering that TDP-43 is a ubiquitous RNA binding protein, I hypothesize that nuclear TDP-43 loss can lead to large transcriptomic changes and may contribute to alterations in LINE elements. For Chapters 3 and 4 of my thesis, I use a novel method of subcellular fractionation and fluorescent activated cell sorting (FACS) from post-mortem ALS/FTD human brain to perform high-throughput sequencing analyses to study neuronal molecular changes. In Chapter 3 of my thesis, I use FACS coupled with RNA-seq on neuronal nuclei with and without

TDP-43 to show that loss of nuclear TDP-43 is associated with large transcriptome changes and increased LINE accessibility. Furthermore, loss of nuclear TDP-43 leads to increased retrotransposition. I also extend this subcellular fractionation-FACS method to study the effects of the *C9orf72* expansion in neuronal nuclei. In Chapter 4, I demonstrate that the *C9orf72* expansion is linked to mild gene expression changes that reflect C9orf72 protein loss and not gain of toxic *C9orf72* RNA. Through my work, I have shown disease mechanisms linked to repetitive DNA elements, in that I propose (1) the *C9orf72* repeat expansion may contribute to disease primarily via a gain of toxic *C9orf72* pathology (2) loss of neuronal nuclear TDP-43 may be associated with increase retrotransposon activity which may contribute to disease. Overall, my work has broadened the field of neurodegeneration in my implementation of cell-type specific molecular analyses on post-mortem brain of ALS/FTD patients to identify disease mechanisms with the intent of discovering new therapeutics and biomarkers that can be extended into the clinic.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

Parts of this introduction has been published in the following manuscript: "RNA metabolism in neurodegenerative disease" by ***Liu E.Y.****, Cali C.P., and Lee E.B. in Disease Models and Mechanisms. doi: 10.1242/dmm.028613*

PREFACE:

Aging-related neurodegenerative diseases, such as Alzheimer's disease (AD), Parkinson's disease (PD), frontotemporal degeneration (FTD) and amyotrophic lateral sclerosis (ALS) among others are ceaselessly progressive and uniformly fatal neurological diseases that are characterized by irreversible neuron loss and gliosis. While dementia prevalence as a percentage of the elderly has declined in developed countries, the absolute number of dementia cases is growing due to the aging of the population (Langa et al., 2017). Thus, it is important for us to develop new methods that will allow us to better understand the basic biological mechanisms that contribute to neurodegeneration.

Brain Heterogeneity Creates Technical Issues for Biological Analysis

The adult human brain is composed of multiple cell types with great complexity that can be subdivided based on function, topology and molecular signatures. Different brain regions, including frontal cortex, parietal cortex, and cerebellum, among others have varying compositions of neurons and non-neuronal cells including glial (astrocytes, oligodendrocytes, and microglia) and vascular cells (endothelial cells and pericytes). For example, in the whole brain, there is an equal ratio of neurons to non-neuronal cells but upon regional specification, there is a clear difference in the non-neuron to neuron ratio. In the cortex, there is a 3.76:1 non-neuron:neuron ratio, whereas in the cerebellum, there is a 0.23:1 non-neuron:neuron ratio (Azevedo et al.,

2009). Furthermore, the cellular diversity within brain regions can affect the molecular signature, including, but not limited to: the transcriptome, epigenetic signature and proteome. Indeed, transcriptional and proteomic differences in varying cell types within the whole brain have been observed, indicating the general complexity of the brain (Hawrylycz et al., 2012; K. Sharma et al., 2015). Through these analyses, it is clear that the major brain cell classes exhibit different molecular signatures, implying that cell type variations reflect different gene expression or epigenetic features between neuronal and non-neuronal cells (Darmanis et al., 2015; K. Sharma et al., 2015). The molecular heterogeneity of the brain makes whole brain molecular analyses difficult to interpret and undermines the complexity and contribution of each cell type in the brain.

Due to the great heterogeneity within the brain, there is a need to develop techniques that will isolate relevant cell types for basic cell biology or disease research. These techniques rely on using cell-type specific markers to either enrich for cells or deplete cells. Some of these markers rely on inherent properties of the cell including cellular adhesion, size, density, morphology and cell surface markers. An additional problem with brain cells lies in its inherent interconnected morphology of axons, dendrites, and glial processes connecting to neighboring cells, which makes it difficult to keep an intact brain cell. Previous efforts to isolate specific neuronal and non-neuronal cell types have relied on laser capture microdissection (LCM) (Chung et al., 2005; Y. M. Jiang et al., 2005; Kamme et al., 2003) where a precise laser is used to dissect out desired cell types from sections of fixed or frozen tissue. While this method allows for specific isolation of cells of interest and can maintain an intact cell, problems that can occur are low yields, costly equipment and compromised RNA quality if the procedure takes too long or if fixation compromises RNA quality (Datta et al., 2015; Fend & Raffeld, 2000). Other efforts including immunopanning have been used to define the transcriptional landscape of the human brain (Y. Zhang et al., 2014; Y. Zhang et al., 2016). Immunopanning relies on an enzymatic digestion to produce a single cell suspension that is passed over a series of antibody-coated dishes and final removal of purified cells at the end. However, this method has some technical difficulties in that cellular dissociation and immunopanning requires time to ensure cellular viability

and may result in low yield if the proper reagents are not used (Barres, 2014). Moreover, dissociation of brain tissue into single cell suspensions can only reliably be done on fresh tissue, precluding the use of archived fixed or frozen specimens. Single cell sequencing of individual cell types have also been implemented to study neuron and non-neuronal cell types, whereby this method relies on single cell sorting after mechanical homogenization (Chiu et al., 2014; Dulken, Leeman, Boutet, Hebestreit, & Brunet, 2017). While single cell sequencing techniques have proven very powerful, the resulting analysis from single cells can be transcriptional noise, technical artifacts or stochastic bursts which may not have biological relevance to the cell types in question (Grun, Kester, & van Oudenaarden, 2014; J. K. Kim, Kolodziejczyk, Ilicic, Teichmann, & Marioni, 2015). Moreover, practical considerations, in particular the cost of single cell analysis, can be prohibitive. As a consequence, there is a need to develop better techniques to study cell type specific effects between different neuronal and non-neuronal cells. This is particularly important for neurodegenerative diseases as the degenerating brain represents cell types that are undergoing disease processes, which naturally make analysis interpretation difficult.

Post-mortem neurodegenerative disease brain reveals a snapshot of a patient brain undergoing disease. This means that the brain sample represents (1) cells that are relatively healthy, (2) cells that are undergoing the disease process and exhibit pathology and (3) reactive glial cells. A longstanding and unsolved question about neurodegeneration is why certain cell types are selectively vulnerable to degenerate (Coyle & Puttfarcken, 1993). For example, in the case of amyotrophic lateral sclerosis (ALS), a neurodegenerative disease that is characterized as loss of upper and lower motor neurons, it has been shown that motor neurons are particularly sensitive to excessive stimulation of glutamate receptors such that excitotoxicity may lead to selective motor neuron death (Van Den Bosch, Van Damme, Bogaert, & Robberecht, 2006). The molecular phenotype for these cells should be considered in trying to understand selective vulnerability. Consequently, it is ever more important to recognize that the heterogeneous nature of the brain needs to be de-convoluted using cell-type specific analyses.

Molecular Changes in Neurodegenerative Diseases

At a basic molecular level, there are multiple levels of genetic regulation. The central dogma is that DNA encodes the genetic signature in cells, which gets transcribed into RNA and then subsequently gets translated into protein. However, modes of gene regulation are becoming increasingly complex. At the DNA level, higher order chromatin structure and epigenetic modifications including DNA methylation and chromatin remodeling are a few of the factors that can dictate what DNA sequences will be accessible for transcription. At the post-transcriptional level, micro-RNAs (miRNAs), long non-coding RNAs (lncRNAs) or RNA binding protein bindings are a few factors that can dictate how RNA will be transcribed and spliced to produce mRNA. Once the mRNA is produced, mRNA will be recruited to ribosomes to create protein. For the purpose of this introduction, some of the pathways, which affect RNA and DNA will be discussed in the context of neurodegeneration.

Although much emphasis has been placed on the role of protein aggregates in neurodegenerative diseases, multiple lines of evidence also converge on altered DNA and RNA processing as a contributing factor in the pathogenesis of these diseases (Anderson & Ivanov, 2014; Belzil, Gendron, & Petrucelli, 2013; Bentmann, Haass, & Dormann, 2013; Halliday et al., 2012; S. C. Ling, Polymenidou, & Cleveland, 2013). Defects at all levels of gene regulation, from changes in the epigenome to defects in RNA synthesis, processing, function and degradation, are associated with disease-specific alterations. Given that these basic processes are essential for normal and properly regulated gene expression, it is increasingly clear that aberrations in these processes can contribute to disease. Here, I highlight several key themes that explain how (1) DNA methylation and chromatin remodeling and (2) different classes of RNAs or RBPs can impair gene regulation. I also highlight specific examples with evidence to show that DNA and RNA alterations are critical features of neurodegeneration.

Epigenetic Modifications in Neurodegeneration

The epigenome, broadly defined as a combination of DNA methylation, chromatin remodeling, and posttranslational modifications of histone proteins among others, serves to regulate gene expression at the DNA level. For the purpose of this introduction, DNA methylation and chromatin remodeling will be covered. Chromatin structure is composed of the nucleosome, consisting of 147 base pairs of DNA wrapped around a histone octamer. The packaging of this structure influences all genomic processes, including DNA accessibility for transcription, recombination and DNA repair (Consortium, 2012). Importantly, the DNA accessibility (open vs closed) can influence whether genes are activated or repressed. Accessible chromatin is known as euchromatin (Fig 1.1A) and highly compacted, transcriptionally silent chromatin is known as heterochromatin (Fig 1.1B). In general, heterochromatic regions tend to be rich in DNA methylation, H3K9 methylation marks and heterochromatin protein 1 (HP1) binding, indicative of gene repression; euchromatic regions are enriched for H3K4 acetylation and methylation, typically associated with gene activation (Jenuwein & Allis, 2001) (Fig 1.1A-B).

DNA methylation is a covalent modification whereby a methyl group is deposited predominantly on cytosine-guanine dinucleotides by DNA methyltransferases (DNMTs). This modification can also be removed by ten-eleven translocation (TET) enzymes, which convert the methylcytosine into hydroxymethylcytosine and finally into formylcytosine and carboxycytosine (Ito et al., 2010; Tahiliani et al., 2009). During DNA repair and cell division, these products can be actively and passively removed. CpGs can be found within promoter regions, repetitive elements or gene bodies. Regions that have more than 60% CpG dinucleotides within a span of 200 base pairs are considered CpG islands, which can be found within gene bodies or promoter regions. The presence of 5-methylcytosine within CpG islands of promoters or repetitive elements is associated with gene silencing and may be used as a defensive mechanism to silence potentially toxic repetitive elements (Deaton & Bird, 2011). On the other hand, CpG methylation within gene bodies has been associated with increased gene expression (X. Yang et al., 2014). Thus, the genomic context of methyl group deposition can impact gene expression as well.

It is thought that CpG methylation promotes a conformational change of DNA structure, which leads to transcriptional repression (Clouaire & Stancheva, 2008) (Fig 1.1B). However, other factors, including transcription factor binding ability or histone modifications around CpG islands can influence methylation status too. For example, removing a transcription factor binding site, Sp1, within the CpG island promoter of mouse gene adenine phosphoribosyltransferase was able to promote de novo methylation of the previously unmethylated promoter region (Macleod, Charlton, Mullins, & Bird, 1994). It is also thought that promoter silencing is modulated by heterochromatin assembly around promoter regions through histone modifiers (i.e H3K9 methyltransferases). These marks can promote DNMTs to deposit methyl groups onto CpGs, thereby leading to promoter methylation (Ayyanathan et al., 2003). These features further highlight the context specificity of DNA methylation.

Aberrations in DNA methylation have been observed in many neurological diseases, including Rett syndrome and some trinucleotide repeat expansion diseases. Rett syndrome is a rare genetic disorder that preferentially affects developing girls and can present as language loss, hand and gait defects or stereotypic hand movements. A mutation in methyl-CpG binding protein (MeCP2), a protein that binds to methyl groups within the genome, causes this disease, highlighting a very important role of DNA methylation regulation in development (Amir et al., 1999; Chahrour et al., 2008). Another set of diseases associated with DNA methylation are microsatellite repeat expansion disorders. They are a class of neurodegenerative diseases caused by repetitive DNA elements that form long expansions within gene bodies or in untranslated regions. DNA methylation has also been observed in two repeat expansions diseases including Fragile X syndrome and Friedreich's ataxia. Fragile X syndrome is the most commonly inherited form of intellectual disability (Gustavson, Blomquist, & Holmgren, 1986; Webb, Thake, & Todd, 1986). The disease is caused by more than 200 CGG repeats in the 5' UTR of the *FMR1* gene (Kremer et al., 1991). In these patients, the CGG repeats and upstream promoter are methylated and this leads to downregulation of FMR protein (FMRP), which is involved in synaptic transmission and mRNA transport (Pieretti et al., 1991; Vincent et al., 1991)

(Fig 1.1C). This downregulation of FMRP is thought to cause disease. Friedreich's ataxia is a degenerative disease where patients have diverse clinical features ranging from ataxia, spasticity, and foot deformities (Durr et al., 1996). Importantly, patients have GAA repeats in the first intron of frataxin (*FXN*), which causes disease (Campuzano et al., 1996). DNA methylation upstream of the repeat was inversely correlated with *FXN* mRNA expression and *FXN* mRNA expression is inversely correlated disease severity, in post-mortem tissue from Friedreich's ataxia patients (Al-Mahdawi et al., 2008; Evans-Galea et al., 2012a). In the context of repeat expansion diseases, it has been thought that methylation of the repeat expansion or surrounding regions would contribute to disease severity.

DNA methylation has also been closely linked to neurodegeneration in AD. AD is the most common form of dementia and is characterized by progressive memory loss, impaired cognitive function, and the inability to perform daily tasks. Pathologically, AD is defined by the presence of extracellular amyloid-beta (Aβ) plaques and of intracellular hyper-phosphorylated neurofibrillary tangles composed of tau (Hardy & Selkoe, 2002). Indeed, two independent reports used epigenome-wide association studies in Alzheimer's disease and non-diseased post mortem brains to identify methylated loci. Both groups identified a hypermethylated locus in the gene ankyrin 1 (*ANK1*) (De Jager et al., 2014; Lunnon et al., 2014). *ANK1* is an adapter protein that is involved in the attachment of integral membrane proteins to the cytoskeleton (Kontrogianni-Konstantopoulos & Bloch, 2003). This differentially methylated region was associated with AD neuropathology in the entorhinal cortex (Lunnon et al., 2014), typically the start of pathological spread in Alzheimer's disease (Braak, Alafuzoff, Arzberger, Kretzschmar, & Del Tredici, 2006). Furthermore, 7 hypermethylated loci (*ANK1, CDH23, DIP2A, RHBDF2, RPL13, SERPINF1, SERPINF2*) identified in Alzheimer's disease patients have been associated with aberrations in gene expression. Six of these 7 genes, including *ANK1*, are connected via a network of known AD susceptibility genes, suggesting that hypermethylation of AD associated genes may be involved in disease onset (De Jager et al., 2014). Consequently, these studies show that aberrant

DNA methylation is also observed in AD patients and the correlation between hypermethylation and increased AD pathology suggest that hypermethylation may be deleterious.

The heterogeneous composition of human brain may complicate the methylation interpretation due to the potential that there are cell-type specific differences in methylation between neurons and non-neurons. Using published data from bisulfite sequencing and an orthogonal method of pyrosequencing on neuronal and non-neuronal cells from brain, increased average CpG methylation in neuronal vs non-neuronal cells from mouse cortex is observed (Kessler et al., 2016). Differentially methylated sites between neurons and non-neurons from post-mortem brain are also enriched in predicted enhancers and distal regions from the transcription start site (Kozlenkov et al., 2014). In the context of the degenerating brain, reactive gliosis can influence the composition of neurons by altering synaptic interactions or neural circuit functions (Burda & Sofroniew, 2014). In light of the cell-type specific methylation differences between neurons and non-neurons, there is a possibility that methylation changes observed in AD patients only occur in a subset of cells (i.e. pathologic neurons). Methylation changes that may be detected are an average of methylation from unaffected and affected cells. It is possible that after factoring such differences, there may be more differentially methylated loci that can be identified and reveal more genes that contribute to disease pathogenesis. Thus, there appears to be a role for DNA methylation in AD pathogenesis but the cell type specific methylation effects should be considered for further interpretation of how methylation status can affect disease.

Heterochromatic regions can include repetitive elements that are silenced, which prevents them from recombining (Saksouk, Simboeck, & Dejardin, 2015) (Fig 1.1B). There is age related loss of epigenetic silencing in yeast, *Drosophila* and even human senescent cells (Larson et al., 2012; Tsurumi & Li, 2012). In *Drosophila*, increasing HP1 protein to promote heterochromatin formation contributes to longevity and improved mobility compared to *Drosophila* where HP1 was reduced (Larson et al., 2012). In patients of Hutchinson-Gilford progeria syndrome, which is a fatal human premature aging disease, there are changes in nuclear

architecture and loss of heterochromatin, evidenced by loss of H3K9me3 marks in induced pluripotent stem cells from patients (G. H. Liu et al., 2011). In Werner's syndrome, which is another premature aging disease characterized by the loss of Werner syndrome ATP-dependent helicase (*WRN*), an embryonic stem cell line where *WRN* was knocked out shows that these cells have larger nuclei and loss of H3K9me3 (W. Zhang et al., 2015). These studies all point to the fact that heterochromatin loss is a main feature associated with aging (Fig 1.1D).

In the context of aging related neurodegenerative disease, chromatin changes have been observed in AD as well. Tau mediated neurodegeneration in AD has been shown to lead to increased transcription in formerly heterochromatinized genes. Using tau transgenic *Drosophila* models, there was reduced total H3K9me2 and HP1 in fly brains, and reducing histone modifying enzymes to promote euchromatin formation exacerbates toxicity. Importantly, in AD post-mortem brain, there is also reduced H3K9me2 and increased expression of previously heterochromatinized genes (Frost, Hemberg, Lewis, & Feany, 2014). Another study has shown that tau interacts with heterochromatic DNA regions and that it can regulate H3K9me3 marks in neuronal nuclei using a murine knockout model of tau. Furthermore, in AD post-mortem brains, pathologic phosphorylated tau is associated with altered distribution of H3K9me3, suggesting that tau can regulate heterochromatin (Mansuroglu et al., 2016) (Fig 1.1D). Similar to the cell-type specific methylation marks associated with neuronal and non-neuronal genes, it would make sense chromatin remodeling can differ in a cell type specific manner too. It is possible that the tissue environment in degenerating brain (i.e. reactive glial cells influencing neighboring neurons) can influence chromatin-modifying activity that may be reflected in the neuronal epigenetic signature. Thus, in order to find causal disease associated chromatin changes, specific cell types should be dissected out from within the degenerating brain. In neurodegenerative disease brain, there are disease-associated changes that can contribute to chromatin remodeling.

Transposable Elements in Aging

Aging associated epigenetic changes leads to chromatin alterations and can activate formerly heterochromatinized regions such as repetitive elements, including transposable elements (TE) (De Cecco et al., 2013; Wood et al., 2016) (Fig 1.1D). TEs are a class of repeat DNA sequences that have the ability to mobilize and insert back into the genome (Kazazian, 2004). TEs can be divided into transposons and retrotransposons. DNA transposons can integrate into the genome directly in a “cut and paste” mechanism. Retrotransposons mobilize by first being transcribed into an RNA intermediate that become a template to be synthesized via reverse transcriptase, at which point the cDNA can integrate back into the genome (Erwin, Marchetto, & Gage, 2014). One predominant retrotransposon is long interspersed nuclear elements (LINE) elements. LINES are the only autonomous transposable element in the human genome (Coufal et al., 2009; Dombroski, Mathias, Nanthakumar, Scott, & Kazazian, 1991; Dombroski, Scott, & Kazazian, 1993; Evrony et al., 2012). LINE elements are at least 6kb long and contain 2 main components: open reading frame (ORF1) which encodes an RNA binding protein and ORF2 which encodes an endonuclease and a reverse transcriptase (Babushok & Kazazian, 2007). Many retrotransposons, including LINEs tend to be silenced by epigenetic modifications. The mechanism of TE silencing has been worked out in embryonic stem cells: the association of DNA-binding Kruppel-associated box domain-containing zinc finger proteins promotes recruitment of KRAB-associated protein 1 (KAP1 or TRIM28) onto the retrotransposon, which in turn triggers heterochromatin formation via histone modifiers, deposition of HP1 and DNMTs (Castro-Diaz et al., 2014; Quenneville et al., 2012; Rowe et al., 2013; Schultz, Ayyanathan, Negorev, Maul, & Rauscher, 2002). Furthermore, MeCP2 can promote sustained DNA methylation and reduced LINE1 (L1) expression (Muotri et al., 2010), thereby supporting a role for DNA methylation in suppressing L1 elements.

TEs have been implicated in aging and neurodegeneration (Maxwell, Burhans, & Curcio, 2011). Loss of TE repression leads to increases in retrotransposon levels and DNA damage in an age dependent manner in *Drosophila* (Wood et al., 2016). Several retrotransposons and LINE-

like elements were found to be elevated in *Drosophila* brain with age. This expression contributes to memory decline and age-dependent death (W. Li et al., 2013; Wood et al., 2016). In fact, L1 expression is increased in aging mouse brain, suggesting that aging is closely linked to L1 derepression (Van Meter et al., 2014). Increased L1 retrotransposition has also been observed in neuronal progenitor cells (NPC) derived from Rett syndrome patients carrying a MeCP2 mutation (Muotri et al., 2010), suggesting that MeCP2 is involved in suppressing L1 activation. TE reactivation can also be observed in neurodegeneration, as in the case of Fragile X tremor ataxia syndrome. In Fragile X tremor ataxia syndrome (FXTAS), patients have 50-200 CGG repeats in the *FMR1* gene and experience adult-onset symptoms similar to that of Fragile X patients (P. J. Hagerman & Hagerman, 2004). The CGG repeats are not hypermethylated in these patients and it is thought that transcription of the expanded *FMR1* allele is toxic (Jin et al., 2003). In a fly model of FXTAS, the CGG repeats can activate select retrotransposons that contribute to neurodegeneration. A RNA binding protein, heterogeneous nucleoparticle protein A2/B1, is sequestered by CGG repeats, which prevents its normal interactions with HP1. This leads to heterochromatin loss (Tan et al., 2012). Furthermore, TE reactivation has also been implicated in ALS such that endogenous retroviruses have been found to be activated and elevated in ALS patient brain and blood (Bowen et al., 2016; W. Li et al., 2015; McCormick, Brown, Cudkowicz, Al-Chalabi, & Garson, 2008; Steele et al., 2005). Thus, retrotranposons activation has been strongly implicated both in aging and neurodegeneration.

There have been interesting findings related to modulating retroviral activity in ALS. There is a small percentage of human immunodeficient virus (HIV) patients who exhibit ALS-like symptoms at an estimated frequency of 3.5 per 1000 HIV infected patients (Moulignier, Moulonguet, Pialoux, & Rozenbaum, 2001). While their clinical presentation may be similar to ALS patients, this group tends to develop disease much younger and have faster disease progression than typical ALS patients (Alfahad & Nath, 2013). According to the El Escorial criteria, a scoring guide for ALS diagnosis, among HIV patients who may exhibit ALS-like symptoms, some have clinically definite ALS (41%), probable ALS (31%) or suspected (28%)

(Alfahad & Nath, 2013). What is rather interesting is the fact that HIV does not infect neurons but infects microglial cells in the CNS, suggesting that motor neuron death may be a secondary consequence of microglial HIV infection (Cosenza, Zhao, Si, & Lee, 2002; Jubelt & Berger, 2001). Interestingly, antiretroviral therapy in these patients has varying clinical benefit, ranging from patients regaining some motor abilities to complete recovery (Bowen et al., 2016; MacGowan, Scelsa, & Waldron, 2001). In a recent study in 5 HIV patients with motor neuron disease, treating them with antiretroviral therapy (ART) either slows the progression of motor defects or leads to complete recovery (Bowen et al., 2016). It is thought that the timing of retroviral therapy initiation relative to symptom onset is important to ensure positive effects. Consequently, there may be some connection between retroviral activation and ALS, implicating some promise in ART for ALS patients.

RNA Binding Proteins Regulate RNA Metabolism

After DNA is transcribed into RNA, RNA binding proteins (RBPs) can play a role in gene regulation at the RNA level. RBPs are essentially required at all levels of RNA processing in both the nucleus and cytoplasm where transcription, splicing, RNA stabilization, and RNA degradation occur (Fig 1.2A). Two notable examples of RBP defects occur in familial and sporadic cases of ALS and FTD. FTD is a neurodegenerative disease associated with neuronal loss in the temporal and frontal cortex. Despite different areas of neuronal atrophy, a common link between ALS and FTD patients is a nuclear RBP called TAR DNA binding protein-43 (TDP-43). Post-mortem brains from human ALS and FTD patients show characteristic mislocalization of TDP-43 from the nucleus into phosphorylated, ubiquitinated cytoplasmic TDP-43 aggregates (Neumann et al., 2006). Indeed, rare disease-causing mutations in *TARDBP*, the gene encoding TDP-43, suggest that TDP-43 dysfunction is sufficient to cause ALS (Gitcho et al., 2008; Kabashi et al., 2008; Sreedharan et al., 2008; Van Deerlin et al., 2008), though the mechanism whereby these mutations cause disease is unclear. TDP-43 functions ubiquitously in RNA processing, including splicing (Buratti et al., 2001; J. P. Ling, Pletnikova, Troncoso, & Wong, 2015; Shiga et al., 2012;

Tollervey et al., 2011), stability (Costessi, Porro, Iaconcig, & Muro, 2014; X. Liu, Li, Zhang, Guo, & Zhan, 2012; Strong et al., 2007), and transport (Alami et al., 2014) (Fig 1.2A). Shortly after the discovery of *TARDBP* mutations, mutations in the gene *FUS/TLS*, which encodes the protein fused in sarcoma/translated in liposarcoma, were identified in a subset of ALS patients, in whom FUS mislocalizes to the cytoplasm (Kwiatkowski et al., 2009; Vance et al., 2009). Similar to TDP-43, FUS interacts with serine arginine (SR) proteins involved in RNA splicing (L. Yang, Embree, Tsai, & Hickstein, 1998) and regulates transcription by recruiting RBPs through non-coding RNAs (X. Wang et al., 2008). An additional example of abnormally localized nuclear RBPs is evident in patients with multisystem proteinopathy (MSP), in whom mutations in the gene encoding heterogenous nucleoriboprotein particle (hnRNP) A1 and A2B1 contribute to disease. MSP is characterized by the progressive degeneration of muscle, brain, motor neurons and bone, which sometimes mainfests as ALS or FTD (H. J. Kim et al., 2013). Mutations in the gene encoding valosin containing protein *(VCP)*, a triple ATPase protein involved in many cellular functions including the endolysosomal degradation, autophagy, and the ubiquitin proteasome system, also causes MSP (Watts et al., 2004).

It is thought that the loss of RBPs through their nuclear mislocalization and/or the toxicity caused by their cytoplasmic aggregation can lead to disease, but their relative contributions to disease remain unclear. Mouse models where antisense oligonucleotides (ASOs) against *Tardbp* depleted TDP-43, show altered global RNA expression affecting 601 genes, and specifically neuronal genes with long introns (Polymenidou et al., 2011). To better understand the targets of these RBPs, TDP-43 and FUS have been immunoprecipitated from mice brains and rat primary neuronal cultures to show that these proteins bind to non-coding RNA sites, namely introns and 3' untranslated regions (UTRs) of thousands of genes (Lagier-Tourenne et al., 2012; Sephton et al., 2011). Some of these genes include ncRNAs, like metastasis associated lung adenocarcinoma transcript 1 (*Malat1*) and nuclear paraspeckle assembly transcript 1 (*Neat1*) (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011; Tollervey et al., 2011). Thus, it is possible that the loss of RBPs influences the processing of these non-coding RNAs and contribute to global RNA

dysregulation (Fig 1.2B). Given that the depletion of these RBPs alters the expression of thousands of genes, it is likely that some, or even all of these changes contribute to disease pathogenesis. While these studies highlight some of the functions of TDP-43 in experimental models, the effect of TDP-43 dysfunction on RNA processing in the human brain is not well characterized.

The formation of cytoplasmic RNA granules that lead to cytoplasmic aggregates has also been proposed to be pathogenic. When cells are stressed, cytoplasmic RNA granules that contain stalled translational complexes are formed. TDP-43, FUS, and other RBPs, such as hnRNPA1 and A2B1, localize to stress granules (H. J. Kim et al., 2013; Y. R. Li, King, Shorter, & Gitler, 2013). Indeed, RBPs with low complexity domains (LCD), such as TDP-43, FUS, hnRNP A1, can phase separate to create dynamic membrane-less organelles or liquid droplets that underlie the transient nature of stress granules (Courchaine, Lu, & Neugebauer, 2016; Molliex et al., 2015). The liquid properties of these organelles are dependent on their constituents. Namely, the intrinsic properties, type, concentration of the RBP, RNAs that the RBPs are bound to and the concentration of the RNA greatly influence these RNA granules (Kroschwald et al., 2015; J. Smith et al., 2016). For example, increasing the concentration of RBPs can reduce the liquid-like properties of these RNA granules, thereby promoting the formation of hydrogels and eventually an insoluble amyloid-like aggregate (Guo & Shorter, 2015; Kato et al., 2012; Y. Lin, Protter, Rosen, & Parker, 2015; Molliex et al., 2015; Xiang et al., 2015). Indeed, disease associated mutations within the LCDs of RBPs can enhance prion-like properties, and accelerate the shift from liquid to solid and disrupt ribonucleoparticle (RNP) granule formation (Murakami et al., 2015). At present, there is little evidence supporting amyloid like fibrillar aggregates within neurons in ALS/FTD, but the relationship between other biophysical assemblies such as hydrogels, and neuronal aggregates is being investigated. Thus, the formation of liquid droplets is another mechanism by which RBP disruption could contribute to disease (Fig 1.2B).

The formation of cytoplasmic aggregates has also been implicated in neurodegenerative disease. Several different rat and mouse models, in which wild-type or mutant *TARDBP* or mutant *FUS* bearing ALS-associated mutations are overexpressed, develop cytoplasmic aggregation and exhibit features of ALS and FTD, including cortical and hippocampal neuronal loss and motor deficits (Huang, Tong, Bi, Zhou, & Xia, 2012; Igaz et al., 2011; Scekic-Zahirovic et al., 2016; A. Sharma et al., 2016; Tsai et al., 2010; Wils et al., 2010; Y. F. Xu et al., 2010). However, none of these models has recapitulated the loss of endogenous nuclear TDP-43 or FUS. Despite the discrepancies between animal models and human pathology, there is *in vitro* evidence to show that cytoplasmic aggregates are toxic. Live tracking of rat primary cortical neurons to assess their survival shows that neurons with cytoplasmic TDP-43 have a greater risk of death, and that this risk depends on the amount of cytoplasmic TDP-43 present (Barmada et al., 2010). This corroborates a yeast model of FUS and TDP-43 overexpression that results in cytoplasmic aggregation of these proteins (Johnson, McCaffery, Lindquist, & Gitler, 2008; Johnson et al., 2009; Sun et al., 2011). TDP-43 toxicity is dependent not only on its RNA-binding ability but also its C terminal end (Elden et al., 2010; Johnson et al., 2009; Voigt et al., 2010), the region where most disease-causing *TARDBP* mutations are found (Gitcho et al., 2008; Kabashi et al., 2008; Sreedharan et al., 2008). While there is no consensus on which mechanism is more toxic, it is likely that both nuclear clearance and cytoplasmic aggregation of RBPs contribute to disease. Indeed, an effort to parse the effects of nuclear TDP-43 loss and TDP-43 cytoplasmic aggregation in a mouse motor neuron-like hybrid cell line (NSC34) shows that both contribute relatively equally to cellular toxicity (Cascella et al., 2016).

At present, potential therapeutic interventions are based on reducing the formation of toxic cytoplasmic aggregates. This is achieved in several ways, for example by: (1) activating the heat shock response; (2) using heat shock protein (Hsp)104 disaggregases; and (3) by modulating the ubiquitin proteasome system and autophagy. HSPs function as molecular chaperones and are involved in protein folding, protein trafficking and in coping with denatured proteins (Lindquist & Craig, 1988). Prior work has shown that the overexpression of a HSP, heat

shock factor 1 (HSF1), in rat primary neuronal cultures overexpressing wildtype TDP-43 prevents TDP-43 cytoplasmic aggregation by interacting with other HSPs to enhance refolding. This reduces toxicity in a human bone marrow neuroblast cell line (SH-SY5Y) overexpressing either wild-type or mutant TDP-43 (H. J. Chen et al., 2016). Components of the heat shock response have also been engineered to remove aggregated proteins. Modified Hsp104 improves its disaggregation capabilities relative to the wildtype Hsp104, and is able to suppress FUS and TDP-43 toxicity in yeast. This provides a potential intervention to eliminate protein aggregates that contribute to toxicity (Jackrel et al., 2014). Enhancing components of the ubiquitin proteasome system or autophagy can also reduce these aggregates. For example, increasing cAMP levels with forskolin in human embryonic kidney cells (293A) cells enhances the ubiquitin proteasome system to clear aggregation-prone proteins, such as FUS and TDP-43, in cells overexpressing either wildtype or mutant forms of both proteins (Lokireddy, Kukushkin, & Goldberg, 2015). Furthermore, two different studies show that using autophagy activators rescues motor dysfunction in transgenic FTD mice and also improves survival of neurons and astrocytes derived from human induced pluripotent stem cells from ALS patients harboring *TARDBP* mutation (Barmada et al., 2014; I. F. Wang, Tsai, & Shen, 2013). Until the toxic mechanism that underlies RBP pathology is uncovered, it is difficult to determine which therapeutic intervention will be the most beneficial to patients.

RBP Sequestration by Microsatellite Repeat Expansions

Microsatellite repeat expansion diseases cause neurodegeneration via different mechanisms (Table 1.1) where one of them is altered RNA processing via sequestration of RBPs by expanded repeat sequences. These expanded sequences sequester RBPs away from their target RNAs. This results in altered RNA splicing and metabolism (Fig 2.3) (Iwahashi et al., 2006; H. Jiang, Mankodi, Swanson, Moxley, & Thornton, 2004; Y. B. Lee et al., 2013). One example of altered RNA metabolism in neurological disease comes from the expansion of a CTG triplet that causes Myotonic Dystrophy (DM) in the 3' UTR of the gene *DMPK* (Brook et al., 1992). This

expansion is transcribed into repeat RNA that forms aggregates, called RNA foci. These aggregates form within the nuclei of patient cells (Davis, McCurrach, Taneja, Singer, & Housman, 1997; Taneja, McCurrach, Schalling, Housman, & Singer, 1995) and recruit a class of RBPs that regulate alternative splicing, called the muscleblind like proteins (MBNLs) (Miller et al., 2000a). By sequestering MBNLs into RNA foci, mutant *DMPK* renders MBNLs unable to regulate splicing and the polyadenylation of hundreds of target genes (R. Batra et al., 2014; Goodwin et al., 2015).

Similarly, repeat expansions associated with ALS/FTD and with Fragile X-associated tremor/ataxia syndrome (FXTAS) also sequester RBPs. In this disease, a CGG repeat-containing RNA transcribed from the *FMR1* gene colocalizes with hnRNPs, MBNL1 and with other RBPs (Iwahashi et al., 2006; Jin et al., 2003; Sofola et al., 2007) thereby altering splicing and microRNA (miRNA) biogenesis in FXTAS patients (Sellier et al., 2013; Sellier et al., 2010). In the most common genetic form of ALS/FTD, a hexanucleotide (G_4C_2) expansion in the first intron of the gene *C9orf72* is bidirectionally transcribed into mutant RNA that forms aggregates in the nucleus (DeJesus-Hernandez et al., 2011; Gendron et al., 2013; Renton et al., 2011; Zu et al., 2013). Current evidence shows that RBPs involved in splicing, such as hnRNPs and the Serine/arginine-rich splicing factors that comprise the spliceosome, are sequestered by mutant repeat-containing RNA (Cooper-Knock et al., 2014; Y. B. Lee et al., 2013). Additionally, the repeat expansion can interfere with transcription of the *C9orf72* gene, resulting in haploinsufficiency of the protein product (Burberry et al., 2016; Ciura et al., 2013a; DeJesus-Hernandez et al., 2011). Reduced transcription is in part due to hypermethylation of the mutant *C9orf72* promoter. Hypermethylation is observed in about one third of *C9orf72* mutation carriers and is associated with reduced mutant RNA accumulation and an attenuated clinical phenotype, suggesting that reduced transcription of mutant *C9orf72* is actually protective against disease (E. Y. Liu et al., 2014; McMillan et al., 2015; Russ et al., 2015). Thus, altered RNA metabolism is clearly implicated in neurodegenerative diseases caused by repeat expansions.

Given that repeat expansions cause widespread disruption to RNA metabolism, it will be challenging to target downstream processes for therapeutic intervention. Therefore, the most

promising therapeutic approaches are those that work upstream to reduce the amount of repeat-containing transcripts. Several studies have used ASOs to target *C9orf72* and DM expansions for degradation via an RNAse H-mediated pathway (Donnelly et al., 2013; J. Jiang et al., 2016; Lagier-Tourenne et al., 2013; Sareen et al., 2013; Wheeler et al., 2012). In fact, ASOs are already in clinical trials for the treatment of Huntington's disease (HD) (Kordasiewicz et al., 2012) and have been approved for the treatment of spinal muscular atrophy (SMA) (FDA, 2016, http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm534611.htm). An alternative to targeting the repeat-containing transcript for degradation is to target proteins that are responsible for transcribing the repeat expansion. For instance, knockdown of the transcription elongation factor *SUPT4H1* selectively decreases the repeat containing RNA in *C9orf72* expansion carrier fibroblasts (Kramer et al., 2016). This strategy is attractive because transcription of the repeat is blocked in both the sense and antisense directions (J. Jiang & Cleveland, 2016; Kramer et al., 2016). A final therapeutic approach involves small molecules that target the expanded RNA to prevent their interactions with RBPs. Several of these compounds have been identified but are still in early stages of development (Disney et al., 2012; Luu et al., 2016; Su et al., 2014). While the repeat-expansion disorders offer a clear therapeutic target, reversing alterations in RNA metabolism will be challenging in neurodegenerative diseases that lack a clear genetic etiology.

Defining Molecular Signature Associated with Neurodegenerative Disease

Because there are many molecular changes associated with neurodegenerative disease and each change can add additional stressors to the cell, there is a need to understand how each modification either on the DNA, RNA or protein level contributes to disease. Currently, molecular analyses have focused on using whole brain from neurodegenerative disease patients, which reflects the degenerating tissue's altered cellular composition (i.e. neuron loss and subsequent increase in glial cells). In fact, when comparing whole cortex and purified cell types from an AD mouse model, the amyloid driven changes in whole tissue was primarily due to microglial cells. In

fact, when the same analysis is extended to identify transcriptome changes linked to an ALS mouse model, the results were similar to that found in AD mouse brain. Transcriptome changes highlight microglial alterations, suggesting that these are not disease specific processes linked to ALS. Using whole brain as opposed to purified cell types may reflect changes linked to reactive glial cells present in all neurodegenerative diseases (Srinivasan et al., 2016). This bolsters the fact that analyzing whole diseased brain is not reflective of distinct disease pathologies and stresses the importance of using cell type specific isolation methods to study disease associated molecular changes.

Consequently, in analyses described in my thesis, there is strong emphasis to use gray matter to enrich for neurons or use methods to isolate cell types of interest to analyze disease specific molecular changes in neurons. In doing so, there is a hope to identify distinct molecular pathological signatures that will reveal novel disease mechanisms with the eventual desire to define therapeutic targets or diagnostic biomarkers that can be used in the clinic.

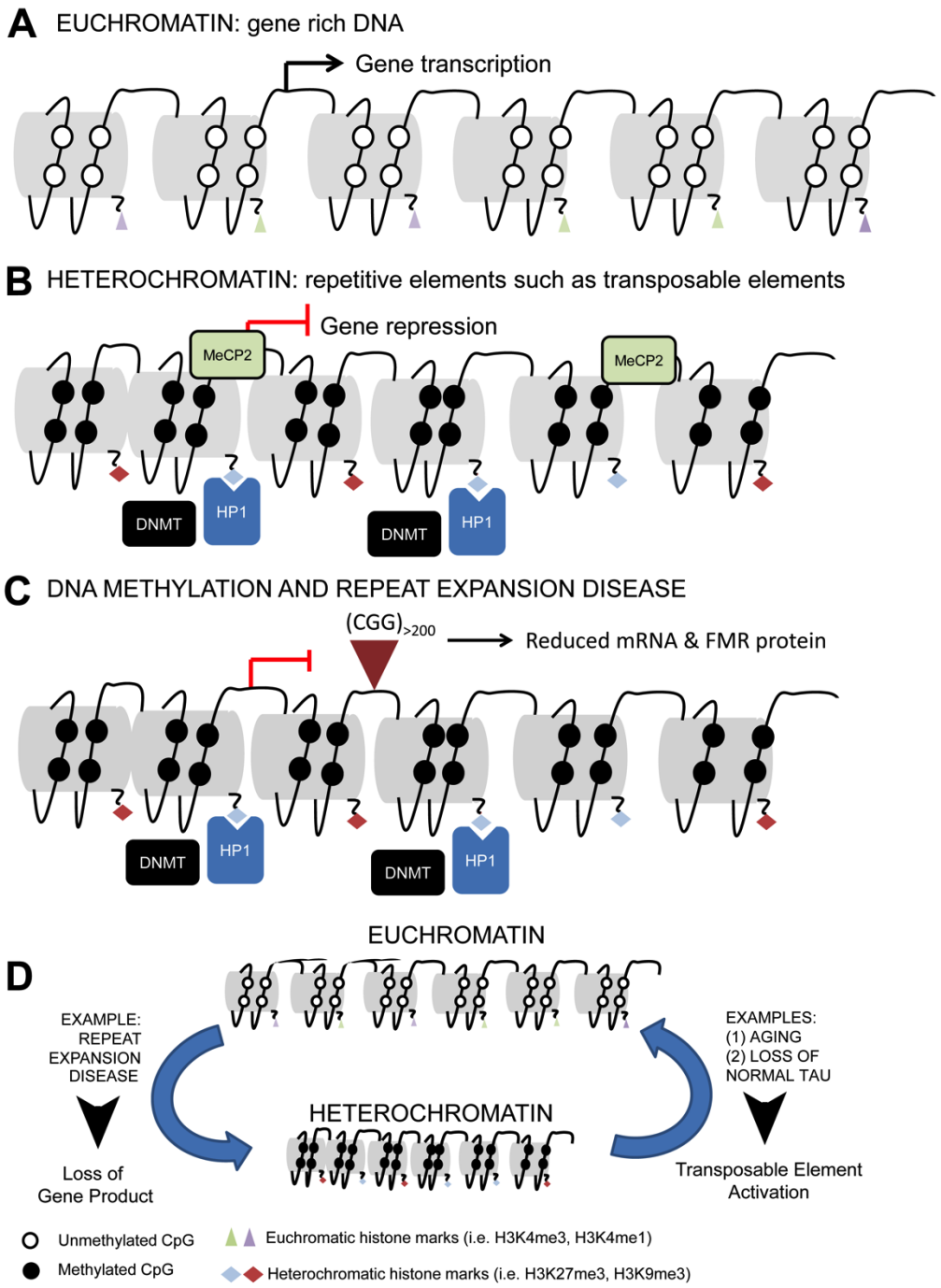

Fig 1.1: **Epigenetic Modifications in Disease.**
(A) Euchromatic regions are typically associated with gene activation, unmethylated CpG island regions (white circles), histone marks associated with gene activation (colored triangles) and tend to be actively transcribed regions of DNA.
(B) Heterochromatic regions are typically associated with silenced DNA, methylated CpG islands (black circles) and repressive histone marks (colored diamonds). Repressive histone marks can recruit heterochromatin binding protein 1 (HP1), which in turn interact with DNA methyltransferases (DNMTs) and methyl-CpG binding protein 2 (MeCP2) to promote stable

silencing of DNA via DNA methylation. Heterochromatic regions tend to be gene poor regions and typically contain repetitive elements including transposable elements.
(C) In Fragile X syndrome, more than 200 CGG repeats within the first exon of *FMR1* gene leads to CpG hypermethylation of the *FMR1* promoter region and the CGG repeat. This is also associated with repressive histone marks and subsequent recruitment of HP1 and DNMTs. Reduced mRNA and FMR protein expression causes disease.
(D) Chromatin changes can occur as a response to cellular stresses. For example, aging or tau hyperphosphorylation as seen in Alzheimer's disease is associated with heterochromatin loss. Loss of formerly heterochromatinized regions, including transposable elements, may lead to increased retrotransposition. Formation of heterochromatin as seen in some repeat expansion diseases can reduce protein expression of essential genes, thereby leading to toxicity.

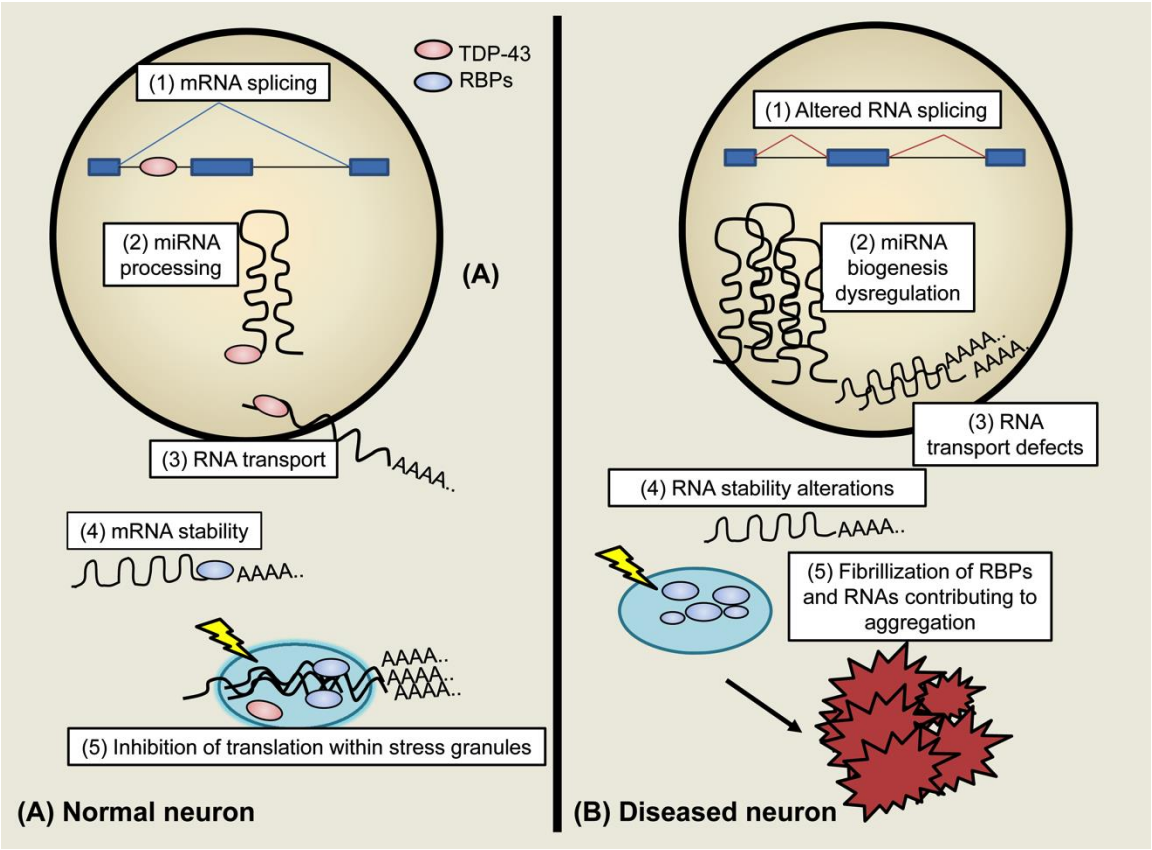

Fig 1.2: **RBPs in healthy and diseased neurons**
(A) RBPs such as TDP-43, FUS/TLS and other hnRNPs are essential for RNA metabolism. Shown here is how TDP-43 can contribute to neurodegeneration. In normal neurons, TDP-43 can: (1) bind RNA in the nucleus to regulate mRNA splicing; (2) bind pri-miRNAs with Drosha to cleave pri-miRNAs into pre-miRNAs; (3) transport RNAs into the cytoplasm; (4) bind 3' UTRs of transcripts to alter mRNA stability; and (5) localize to RNPs (stress granules), filled with RBPs and mRNAs, which halt translation.
(B) In diseased neurons (right panel), nuclear loss of TDP-43 and subsequent cytoplasmic aggregation can alter RNA metabolism. TDP-43 loss can result in: (1) altered RNA splicing, leading to altered gene regulation; (2) altered miRNA biogenesis, leading to increased pri-miRNA and reduced pre-miRNA levels; (3) transcript nuclear retention; and (4) altered transcript stability. Current models cannot recapitulate loss of endogenous TDP-43 and cytoplasmic aggregation concurrently. (5) TDP-43 and other RBPs can also reportedly form liquid droplets that might promote aggregate formation in an RNA- and protein-dependent manner. These structures can trap translational machinery proteins or other RBPs, thereby inhibiting their normal function.

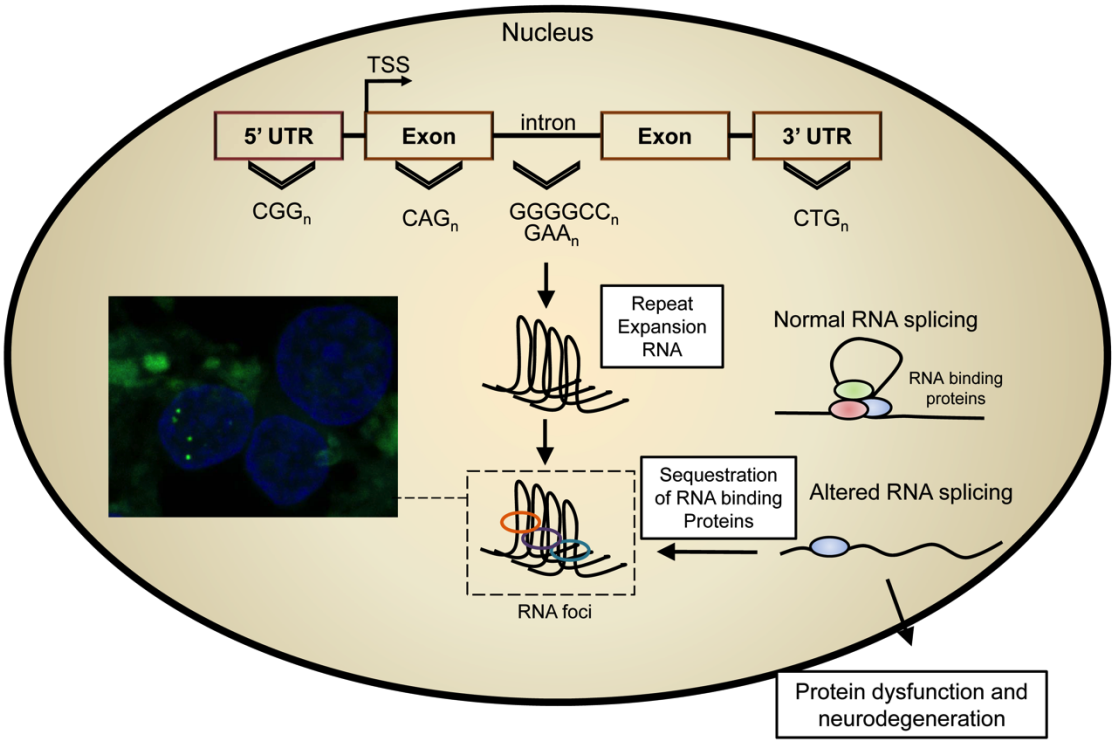

Fig 1.3: **Microsatellite repeats and neurodegenerative disease**
Repeat expansions can occur within the 5' UTR, exon or 3'UTR of a gene. RNA transcripts containing these repeat sequences can aggregate in the nucleus, where they form RNA foci (left image). Depicted here is an image of RNA foci labeled with a fluorescent probe against the G4C2 repeat sequence in post-mortem cerebellar granule cells from an ALS/FTD patient bearing the *C9orf72* expansion. The repeat-containing RNAs recruit RBPs, such as the muscleblind like proteins, hnRNPs and the serine/arginine-rich splicing factors, into RNA foci. The lack of functional RBPs alter RNA splicing and mRNA processing, causing downstream protein dysfunction and eventual neurodegeneration.

Gene Name	Disorder	Repeat Sequence	Location of Expansion	Possible mechanism	References
C9orf72	Amyotrophic Lateral Sclerosis; Frontal Temporal Degeneration	GGGGCC	Intron	Toxic repeat RNA and repeat protein	(Ash et al., 2013; DeJesus-Hernandez et al., 2011; Y. B. Lee et al., 2013)
FXN	Friedreich ataxia	GAA	Intron	Loss of normal gene product	(Campuzano et al., 1996; Grabczyk & Usdin, 2000)
DMPK	Myotonic Dystrophy Type 1	CTG	3' UTR	Toxic repeat RNA and repeat protein	(Brook et al., 1992; H. Jiang et al., 2004; Zu et al., 2011)
CNBP (ZNF9)	Myotonic Dystrophy Type 2	CCTG	Intron	Toxic repeat RNA	(Liquori et al., 2001)
FMR1	Fragile X-associated tremor/ataxia syndrome	CGG (50-200)	5' UTR	Toxic repeat RNA and repeat protein	(Oh et al., 2015; Sofola et al., 2007)
FMR1	Fragile X Syndrome	CGG (>200)	5' UTR	Loss of normal gene product	(Huber, Gallagher, Warren, & Bear, 2002; Pieretti et al., 1991)
HTT	Huntington disease	CAG	Exon	Toxic repeat protein	(MacDonald et al., 1993; Mangiarini et al., 1996)
ATXN8/SCA8	Spinocerebellar ataxia 8	CAG/CTG	3' UTR/Exon	Toxic repeat RNA and repeat protein	(Daughters et al., 2009; Moseley et al., 2006)
ATXN1-3; ATXN6-7; ATXN-17	Spinocerebellar ataxia	CAG	Exon	Toxic repeat RNA and repeat protein	(Durr, 2010; Loureiro, Oliveira, & Silveira, 2016)

Table 1.1: **Microsatellite repeats implicated in neurodegenerative disease.** There are over two dozen human genes that contain repeat expansions. Listed here are the most common expansions that cause neurologic disease, along with the likely mechanism of action.

CHAPTER 2:

C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD

*This chapter has been adapted from the following manuscript: "C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD" by **Liu E.Y.**, Russ J, Wu K, Neal D, Suh E, McNally A.G., Irwin D.J., Van Deerlin V.M., Lee E.B. in Acta Neuropathologica. Cali C.P., and Lee E.B. in Disease Models and Mechanisms. doi: 10.1007/s00401-014-1286-y*

Abstract

Hexanucleotide repeat expansions of *C9orf72* are the most common genetic cause of amyotrophic lateral sclerosis and frontotemporal degeneration. The mutation is associated with reduced *C9orf72* expression and the accumulation of potentially toxic RNA and protein aggregates. CpG methylation is known to protect the genome against unstable DNA elements and to stably silence inappropriate gene expression. Using bisulfite cloning and restriction enzyme based methylation assays on DNA from human brain and peripheral blood, we observed CpG hypermethylation involving the *C9orf72* promoter in cis to the repeat expansion mutation in approximately one third of *C9orf72* repeat expansion mutation carriers. Promoter hypermethylation of mutant *C9orf72* was associated with transcriptional silencing of *C9orf72* in patient-derived lymphoblast cell lines, resulting in reduced accumulation of intronic *C9orf72* RNA and reduced numbers of RNA foci. Furthermore, demethylation of mutant *C9orf72* with 5-aza-deoxycytidine resulted in increased vulnerability of mutant cells to oxidative and autophagic stress. Promoter hypermethylation of repeat expansion carriers was also associated with reduced accumulation of RNA foci and dipeptide repeat protein aggregates in human brains. These results indicate that *C9orf72* promoter hypermethylation prevents downstream molecular aberrations associated with the hexanucleotide repeat expansion, suggesting that epigenetic silencing of the mutant *C9orf72* allele may represent a protective counter-regulatory response to hexanucleotide repeat expansion.

Introduction:

Amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) exhibit overlapping clinical, pathologic and genetic features (Lee, Lee, & Trojanowski, 2012; Mackenzie, Frick, & Neumann, 2014). ALS is characterized by weakness and spasticity due to loss of motor neurons while FTD is characterized by behavioral and language dysfunction due to degeneration of the frontal and temporal lobes. However, many individuals exhibit clinical features of both ALS and FTD, and affected CNS regions exhibit similar neuropathologic changes (Lee et al., 2012; Mackenzie et al., 2014). An intronic GGGGCC hexanucleotide repeat expansion in *C9orf72* is the most common genetic cause of ALS and FTD (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The *C9orf72* mutation is associated with highly variable clinical phenotypes (Boeve et al., 2012; Chio et al., 2012; Cooper-Knock, Hewitt, et al., 2012; M. Harms et al., 2013; Hsiung et al., 2012; Khan et al., 2012; Mahoney et al., 2012; Murray et al., 2013; Simon-Sanchez et al., 2012; Snowden et al., 2012). While the basis for this heterogeneity is largely unknown, this suggests that there are endogenous mechanisms which modulate *C9orf72*-dependent disease pathways.

Mounting evidence suggests that repeat expansion of *C9orf72* results in a toxic gain of function associated with the formation of RNA foci containing repeat RNA (DeJesus-Hernandez et al., 2011) or the accumulation of repeat associated non-ATG mediated translation (RANT) dipeptide aggregates (Ash et al., 2013; Mori et al., 2013). The repeat expansion is associated with increased vulnerability to cellular stressors (Almeida et al., 2013; Donnelly et al., 2013; Sareen et al., 2013) and exogenous expression of the hexanucleotide repeat leads to overt toxicity (Y.-B. Lee et al., 2013). However, the endogenous mechanisms which modulate the pathogenesis of *C9orf72*-associated disease are not well understood. Antisense oligonucleotides that target mutant RNA for post-transcriptional degradation have recently been shown to mitigate the molecular signatures of disease in experimental models (Donnelly et al., 2013; Lagier-Tourenne et al., 2013; Sareen et al., 2013), suggesting that mechanisms which regulate mutant RNA expression may modulate disease pathogenesis.

Repeat expansions have been associated with epigenetic silencing, most notably in the case of trinucleotide repeat mutations of *FMR1* in Fragile X syndrome where promoter hypermethylation is linked to a pathogenic loss of function (Dion & Wilson, 2009). DNA methylation is an epigenetic modification which protects the genome against deleterious repeat DNA elements and regulates imprinted or developmentally-timed gene expression (Straussman et al., 2009). Recent studies have demonstrated that mutation carriers exhibit *C9orf72* promoter hypermethylation (Xi et al., 2013) in association with repressive histone marks (Belzil, Bauer, et al., 2013) that silence gene expression. However, the downstream effects of epigenetic silencing of *C9orf72* have not yet been demonstrated. We hypothesized that endogenous transcriptional gene silencing alters downstream disease pathways by virtue of modulating mutant RNA expression and that understanding the relationship between epigenetic silencing of *C9orf72* and repeat expansion-associated toxicity can distinguish whether the *C9orf72* mutation causes disease via a loss or gain of function mechanism.

Here, we show that *C9orf72* promoter hypermethylation is found in a subset of repeat expansion cases and appears to protect against mutant *C9orf72* mediated toxicity and pathology. Using human tissue, we determined that *C9orf72* hypermethylation is monoallelic involving a region upstream of the hexanucleotide repeat only in repeat expansion mutation carriers. *C9orf72* promoter methylation resulted in transcriptional silencing, inhibiting the accumulation of mutant intronic RNA, reducing vulnerability to oxidative and autophagic stress, decreasing the number of RNA foci, and inhibiting the accumulation of RANT pathology. Collectively, these results support the hypothesis that *C9orf72* mutations cause disease through a toxic gain of function, and indicate that *C9orf72* promoter hypermethylation may represent an endogenous response that modulates disease via inhibition of mutant *C9orf72*.

Materials and Methods:

Patient-Derived Materials

Human autopsy tissue and peripheral DNA was obtained from the University of Pennsylvania Center for Neurodegenerative Disease Research biorepository (Toledo et al., 2013). A summary of cases is available in supplemental materials. Lymphoblast cell lines ND16183, ND11836, ND14442 and ND10966 were obtained from the Coriell NINDS Repository (Camden, NJ).

Nucleic Acid Extraction

Total RNA was extracted using Trizol (Life Technologies, Carlsbad, CA). RNA was digested with 1 unit of RQ1 DNase (Promega, Madison, WI) per µg followed by ethanol precipitation. RNA was quantified using the Qubit RNA HS Assay kit (Life Technologies). RNA was reverse transcribed to cDNA using the High Capacity RNA to cDNA kit (Life Technologies). qPCR was done with 2x FastStart SYBR Green Master (Roche Applied Science, Indianapolis, IN) to quantify *C9orf72* RNA and control housekeeping RNAs using primers as listed in the supplemental materials on the StepOne Plus Real-Time PCR Machine (Life Technologies) using the ΔΔCt method. For qPCR measuring *C9orf72* intronic RNA, reactions containing RNA without reverse transcriptase did not amplify. DNA from tissue or cells was extracted using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA).

C9orf72 Promoter Methylation Assay

For quantitative assessment of methylation levels, 100 ng of DNA was digested for 16 hours with 2 units of HhaI (New England Biolabs, Ipswich, MA) followed by heat inactivation. qPCR was done with 2x FastStart SYBR Green Master (Roche) using primers amplifying the differentially methylated *C9orf72* promoter region (see Table 2.2 for primer sequences). The difference in the number of cycles to threshold amplification between digested versus mock

digested DNA was used as a measure of CpG methylation. Mock digested DNA consisted of either undigested DNA or HaeIII digested DNA as a negative control restriction enzyme which does not cut within the qPCR amplicon. To determine the linearity of this assay, a methylated DNA standard was generated by *in vitro* methylating DNA from a non-expanded lymphoblast cell line using M.SssI (New England Biolabs) for 4 hours at 37°C. Methylated DNA was purified by phenol:chloroform:isoamyl alcohol extraction, and different ratios of methylated to unmethylated DNA were subject to the *C9orf72* promoter methylation assay.

To determine if methylation occurs in cis or trans, 100 ng of DNA from 3 *C9orf72* promoter hypermethylated repeat expanded patient cases who were heterozygous for the deletion polymorphism (rs200034037) was digested for 16 hours with 2 units of HhaI and HpaII (New England Biolabs) vs. a no enzyme mock digestion followed by heat inactivation. DNA was amplified by PCR using primers flanking rs200034037 and the HhaI and HpaII cut sites within the *C9orf72* promoter (Table 2.2 for primer sequences). The PCR product was run on a polyacrylamide gel and imaged with ethidium bromide or used for Sanger sequencing.

Bisulfite Cloning

Cerebellar DNA from 4 *C9orf72* expansion carriers and 4 control cases was bisulfite converted using the EpiTect Bisulfite Kit (Qiagen). Bisulfite converted DNA was subjected to PCR to amplify both CpG islands. The entire first CpG island was amplified by standard PCR. The second CpG island was divided into two separate reactions. The first half of the 2nd CpG island was amplified by nested PCR, while the second half of the second CpG island was amplified using standard PCR. Primer sequences are available in the supplemental materials. PCR products were run on an agarose gel and gel purified using QIAquick Gel Extraction kit (Qiagen). Amplified DNA was sub-cloned by ligation into the pGEM-T Easy vector (Promega), transformation into competent *E. coli* and isolation of plasmid DNA from individual bacterial colonies. Five or six individual clones from each case were Sanger sequenced using a T7 promoter sequencing primer.

Hexanucleotide Repeat Methylation Assay

DNA (1 µg) from post-mortem cerebellar tissue was restriction enzyme digested for 4 hours with HpaII (10U), MspI (10U), or MspJI (4U) (New England Biolabs) at 37°C followed by phenol:chloroform:isoamyl alcohol extraction. 100 ng of digested DNA was used for repeat primed-PCR as described previously (Renton et al. 2011). Fragment length analysis was done using the Genetic Analyzer 3130x (Life Technologies) and Peak Scanner software (Life Technologies). To ensure that methylated DNA could be digested by MspJI under these conditions, 1 µg of DNA from repeat expanded or control cerebellum was *in vitro* methylated using M.SssI (New England Biolabs) for 4 hours at 37°C followed by phenol:chloroform:isoamyl alcohol extraction. 1 µg of purified DNA was digested with 2 units of MspJI at 37°C for 4 hours and purified using phenol:chloroform:isoamyl alcohol prior to repeat primed-PCR.

Southern Blotting

Southern blot hybridization was performed as previously described (DeJesus-Hernandez et al., 2011) . Briefly, 5-10 µg genomic DNA was digested with EcoRI and HindIII, denatured at 95°C for 5 minutes, and run on a 0.8% agarose gel at 100V for 4 hours. DNA was transferred to a positively charged nylon membrane (GE Life Sciences, Pittsburg, PA) and crosslinked. A 576 bp digoxigenin (DIG)-labeled probe was amplified using PCR DIG Probe Synthesis Kit. Primers are listed in supplemental materials. The blot was hybridized for 16 hrs at 49°C, washed in 2X SSC with 0.1% SDS at room temperature, and then in 0.1X SSC, 0.1% SDS for 45 min at 71°C twice. Anti-digoxigenin antibody (1:10000, Roche) was used to detect the probe, which was visualized with CSPD (Roche). Blots were visualized with an LAS-3000 Luminescent Image Analyzer (Fujifilm).

5-aza-deoxycytidine Treatment

Lymphoblast cell lines derived from either a repeat expansion carrier (ND14442) or a non-expanded ALS patient (ND16183) were grown in RPMI 1640 with 15% fetal bovine serum and 2mM L-glutamine in 37ºC with 5% CO2. Cells were seeded at a concentration of 300,000 cells/mL in 10 mL in a T25 flask, and 0.5 μM 5-aza-dC (Sigma-Aldrich, St. Louis, MO) was spiked daily into lymphoblast cells for 3 days followed by a 3 day recovery period without 5-aza-dC. RNA and DNA was extracted as described above. To assess toxicity, untreated or 5-aza-dC treated cells were seeded at 300,000 cells/ml and treated 10 μM sodium arsenite or 100 μM chloroquine for 24 hours in RPMI 1640 with 5% FBS and 2mM L-glutamine. After 24 hours, cells and media were collected and measured in triplicate using a lactate dehydrogenase assay (Clontech, Mountain View, CA) to calculate the percent LDH release into the media relative to total LDH from untreated cell pellets.

RNA Flourescent *in situ* Hybridization (FISH)

Lymphoblast cell lines were fixed in 4% paraformaldehyde (PFA) in RNase free PBS for 10 minutes on ice followed by permeabilization with 0.2% Triton X-100 in PBS for 10 minutes on ice. Cells were prehybridized with 40% formamide (Fisher Scientific)/1x SSC at 37°C for 10 minutes. A locked nucleic acid (LNA) probe complementary to the hexanucleotide repeat (FAM-CCCCGGCCCCGGCCCC, batch #613510, Exiqon, Woburn, MA) was denatured at 85°C for 75 seconds prior to incubation with cells in hybridization buffer (40% formamide, 1x SSC, 50mM sodium phosphate, pH=7, 10% RNase free dextran sulfate (Sigma-Aldrich) with 40nM LNA probe) at 66°C for 2 hours. Cells were washed once with 0.1% Tween-20 in 2x SSC at room temperature for 5 minutes, and three times in 0.1x SSC at 65°C for 10 minutes each. Cells were then resuspended in Prolong Gold Antifade Reagent with DAPI (Life Technologies), coverslipped onto glass slides, and imaged on a Leica SPE confocal microscope (Leica Microsystems, Buffalo Grove, IL). RNA foci numbers were scored in 200 cells for each cell line blinded to cell type using a 60x objective.

Frozen cerebellar cortex (100-200mg) was dounce homogenized in 2 mls of 0.25M sucrose with TKM (50mM Tris-HCl, pH7.5, 25 mM potassium chloride and 5mM magnesium chloride). Two volumes of 2.3M sucrose with TKM were mixed with the lysate. Nuclei were pelleted at 10,000xg for 10 minutes at 4°C. Nuclei were fixed in 2% PFA in RNase free PBS for 10 minutes at room temperature and quenched with 0.83M glycine (pH=7.6). Nuclei were pelleted and resuspended in RNase free PBS. Nuclei were pre-hybridized with hybridization buffer (40% formamide, 10mM ribonucleoside vanadyl complex (New England Biolabs), 1x SSC, 50mM sodium phosphate, pH=7, 10% RNase free dextran sulfate) for 10 minutes at 37°C. The LNA probe complementary to the hexanucleotide repeat was denatured at 100°C for 10 minutes in 95% formamide and then hybridized to nuclei for 12-16 hours at 37°C at 40nM. Nuclei were washed and imaged as described above for LCLs. RNA foci numbers were scored in at least 100 nuclei for each case using a 60x objective.

RANT Immunohistochemistry

Polyclonal anti-$(GA)_{15}$ and polyclonal anti-$(GP)_{15}$ antibodies were generated by immunizing rabbits with a purified fusion protein containing his-tagged maltose binding protein (MBP) with $(GA)_{15}$ or $(GP)_{15}$ at the C-terminus. The GA or GP repeats fragment was engineered by using the primers as described in the supplemental materials (Integrated DNA Technologies). The annealed fragment was ligated into the pDB.His.MBP vector (DNASU Plasmid Repository, AZ) at the NdeI-EcoRI cloning sites. The his-tagged MBP-GA_{15} and MBP-GP_{15} fusion proteins were expressed in BL21 bacterial cells upon IPTG induction and purified using Ni-NTA Superflow (Qiagen, CA). $(GR)_{15}$ antibody was generated by immunizing rabbits with an aminohexanoic acid linked synthetic peptide (C-Ahx-(GR15)) conjugated to KLH and affinity purified using SulfoLinkR immobilized immunogen peptide (Thermo Scientific). Cerebellar sections were stained with RANT antibodies using standard ABC methods with microwave antigen retrieval (E. B. Lee et al.,

2013), and at least 300 granular neurons per case were assessed using a 100x objective blinded to methylation status to determine the percentage of neurons with RANT inclusions.

RANT Biochemistry

Frozen cerebellar cortex was homogenized in a series of extraction buffers intermixed with pelleting of insoluble material by ultracentrifugation at 135,000xg for 30 minutes. The extraction buffers included RIPA buffer (5 mls buffer per gram tissue; 50 mM Tris pH 7.3, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40 with protease inhibitors), myelin floatation buffer (5 mls buffer per gram tissue; 10mM Tris pH 7.5, 500 mM NaCl, 2mM EDTA, 1mM DTT, 30% sucrose with protease inhibitors), DNAse buffer containing (Promega, 1 unit of RQ DNAse per 10 mg tissue in 1x reaction buffer, incubated at 37 degrees for 30 minutes) and sarkosyl buffer (5 mls buffer per gram tissue with sonication; 1% sarkosyl, 10mMTris pH 7.5, 500 mM NaCl, 2mM EDTA, 1mM DTT, 10% sucrose with protease inhibitors). The remaining insoluble material was pelleted and sonicated in SDS buffer (1 ml per gram tissue, 2% SDS, 50mM Tris, pH 7.6). Five µl of SDS lysates were dot blotted onto nitrocellulose using a Biorad Bio-Dot apparatus and washed with 300 µl of SDS buffer containing β-mercaptoethanol. Dot blots were blocked with 5% milk, blotted with RANT-specific antibodies and visualized by film using enhanced chemiluminescence.

Statistical Analyses

Two-way ANOVA, one-way ANOVA, t-tests, linear regression and Fisher's exact test were used as described using GraphPad Prism software (GraphPad, San Diego, CA). Post-hoc analyses were Bonferroni corrected. All statistical tests were two-sided.

Diagnosis	MND	Dementia	Gender	Age of Onset	Age of Death	Figure	C9ORF72 Expansion
ALS	Y	N	Female	54	55	2.3, 2.7	Y
ALS	Y	N	Female	55	59	2.1, 2.3, 2.7	Y
ALS	Y	N	Female	73	76	2.3, 2.7	Y
ALS	Y	N	Female		56	2.3, 2.7	Y
ALS	Y	N	Male	56	58	2.3, 2.7	Y
ALS	Y	N	Male	56	58	2.1, 2.3, 2.7	Y
ALS	Y	N	Male	62	64	2.1, 2.3, 2.7	Y
ALS	Y	N	Male		54	2.3, 2.7	Y
ALS	Y	N	Male		55	2.3	Y
ALS	Y	N	Male	51	56	2.1	Y
ALS-dementia	Y	Y	Female	63	70	2.3	Y
ALS-dementia	Y	Y	Male	46	48	2.1, 2.3, 2.7	Y
ALS-dementia	Y	Y	Male	52	54	2.1, 2.3, 2.7	Y
ALS-dementia	Y	Y	Male	55	58	2.3, 2.7	Y
ALS-dementia	Y	Y	Male		63	2.1, 2.3	Y
ALS-dementia	Y	Y	Male		76	2.3, 2.7	Y
FTD	N	Y	Female	59	62	2.3, 2.7	Y
FTD	N	Y	Female	66	76	2.1, 2.3, 2.7	Y
FTD	N	Y	Female		72	2.3, 2.7	Y
FTD	N	Y	Male	47	54	2.3, 2.7	Y
FTD	N	Y	Male	63	76	2.1, 2.3, 2.7	Y
FTD	N	Y	Male	63	67	2.1	Y
AD	N	Y	Female	57	70	2.1, 2.3, 2.7	N
AD	N	Y	Male	72	79	2.3	N
AD	N	Y	Male		83	2.3	N
ALS	Y	N	Female	50	53	2.1, 2.3, 2.7	N

ALS	Y	N	Female	81	82	2.1, 2.3	N
ALS	Y	N	Male	76	77	2.1, 2.3	N
DLB, AD	N	Y	Female	72	84	2.1, 2.3, 2.7	N
FTD	N	Y	Female	67	76	2.1, 2.3	N
normal	N	N	Female	n/a	72	2.3	N
normal	N	N	Female	n/a	78	2.3	N
normal	N	N	Female	n/a	82	2.3	N
normal	N	N	Male	n/a	47	2.3	N
normal	N	N	Male	n/a	74	2.3	N
normal	N	N	Male	n/a	76	2.3	N
PD with MCI	N	Y	Male	83	98	2.1, 2.3, 2.7	N
vascular dementia	N	Y	Male	58	65	2.1, 2.3, 2.7	N

Table 2.1: List of cases

FTD= frontotemporal degeneration with TDP-43 inclusions; AD= Alzheimer's disease; MCI= mild cognitive impairment; DLB= dementia with Lewy bodies; PD= Parkinson's disease

Description	Forward Oligonucleotide	Reverse Oligonucleotide
C9orf72 V2 mRNA (exon 2 to 3)	CGGTGGCGAGTGGATATCTC	TGGGCAAAGAGTCGACATCA
C9orf72 Total mRNA (exon 3 to 4)	ACTGGAATGGGGATCGCAGCA	ACCCTGATCTTCCATTCTCTCTGT GCC
C9orf72 V3 mRNA (exon 1b to 3)	GCAAGAGCAGGTGTGGGTTT	TGGGCAAAGAGTCGACATCA
C9orf72 intron (intron 1)	GCAAGAGCAGGTGTGGGTTT	TCTTGTTCACCCTCAGCGAG
C9orf72 SNP allele (sequencing)	GACCAAAAGACGCAAGGTTT	CTTTTACGTGGGCGGAACT
C9orf72 SNP allele (PAGE)	GACCAAAAGACGCAAGGTTT	TTTGTGCTTGGTAGGCAGTG
C9orf72 promoter	CAGTGTGAAAATCATGCTTGAGAGA	TTTGTGCTTGGTAGGCAGTG
C9orf72 CpG Island 1	GGAGTTGTTTTTTATTAGGGTTTGTAG T	TAAACCCACACCTACTCTTACTAA A
C9orf72 CpG Island 2, first half (outer)	GTGAGTGAGGAGGYGGTATTTT	AATCACTTCCTTTAAACAAAT
C9orf72 CpG Island 2, first half (inner)	GTGAGTGAGGAGGYGGTATTTT	AAACCTCTCAATACCCRAAACT
C9orf72 CpG Island 2, second half	AGTTGTTTTTTTTTYGGGGATT	AATCACTTCCTTTAAACAAAT
C9orf72 Southern Blot	TTGCGATGACTTTGCAGGGGACC	CAGCGAGTACTGTGAGAG
GPS mRNA (control)	AAGATGCTGGACGAGATGAAGGA	ACGGTTGCGAATCTGGGTGTA
BACT mRNA (control)	GCCCTGAGGCACTCTTCCA	ATGCCACAGGACTCCATGC
GA Repeat Fragment	TATGGGTGCGGGAGCTGGAGCTGGT GCAGGTGCAGGTGCAGGAGCAGGTG CTGGAGCAGGAGCTGGTGCAGGTGC AGGAGCAGGAGCCGGAGCCTAAG	AATTCTTAGGCTCCGGCTCCTGCT CCTGCACCTGCACCAGCTCCTGC TCCAGCACCTGCTCCTGCACCTG CACCTGCACCAGCTCCAGCTCCC GCACCCA
GP Repeat Fragment	TATGGGTCCTGGACCTGGACCTGGA CCAGGGCCAGGGCCGGGACCTGGA CCTGGACCAGGACCTGGACCAGGGC CAGGACCAGGACCTGGACCATAAG	AATTCTTATGGTCCAGGTCCTGGT CCTGGCCCTGGTCCAGGTCCTGG TCCAGGTCCAGGTCCCGGCCCTG GCCCTGGTCCAGGTCCAGGTCCA GGACCCA

Table 2.2: Primer Sequences

Cell line	Cell Type	Other Type
A549	Epithelial	
HAOEC	Epithelial	
HELAS3	Epithelial	
HEPG2	Epithelial	
HMEPC	Epithelial	
HPIEPC	Epithelial	
HSAVEC	Epithelial	
HUVEC	Epithelial	
MCF7	Epithelial	
NHEK	Epithelial	
AG04450	Fibroblastic	
BJ	Fibroblastic	
HAOAF	Fibroblastic	
HVMF	Fibroblastic	
IMR90	Fibroblastic	
NHDF	Fibroblastic	
CD20	Hematopoietic	
CD34 mobilized	Hematopoietic	
GM12878	Hematopoietic	
K562	Hematopoietic	
MONOCD14	Hematopoietic	
SKNSH	Neuronal	
SKNSH RA treated	Neuronal	
H1HESC	Other	Embryonic stem cell
HCH	Other	Chondrocytes
HFDPC	Other	Hair follicle dermal papilla cells
HMSCAT	Other	Mesenchymal stem cell from adipose
HMSCBM	Other	Mesenchymal stem cell from bone marrow

HMSCUC	Other	Mesenchymal stem cells from umbilical cord
HOB	Other	Osteoblasts
HPCPL	Other	Pericytes
HWP	Other	White preadipocytes
NHEMFM2	Other	Melanocytes
NHEMM2	Other	Melanocytes
SKMC	Other	Striated muscle

Table 2.3: ENCODE CAGE-seq Cell Lines

Results:

Methylation of the *C9orf72* Promoter

Methylation of repeat expansions can reduce gene transcription as exemplified by Fragile X syndrome (O'Donnell & Warren, 2002). To determine whether the *C9orf72* hexanucleotide repeat expansion is methylated, cerebellar genomic DNA from eight repeat expansion carriers and eight nonexpanded controls were digested using three restriction enzymes: (1) MspI that cuts the hexanucleotide repeat irrespective of methylation status, (2) HpaII that cuts the hexanucleotide repeat only if it is not methylated, and (3) MspJI that cuts the hexanucleotide repeat only if it contains methylated CpGs. Digested DNA was then amplified using repeat primed PCR as a qualitative measure of whether the repeat region is methylated. Repeat-primed PCR allows for partial amplification of the expanded hexanucleotide repeat resulting in a characteristic tapering sawtooth pattern, and is used because conventional PCR is unable to amplify repeat expansions due to their size and high GC content (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Non-expanded controls demonstrated the pattern expected of unmethylated DNA in which DNA was amplified after mock and MspJI digestion, and was not amplified after HpaII and MspI digestion (Figure 2.1a). The same pattern was observed for repeat expansion carriers with amplification of the repeat expansion after MspJI digestion but no amplification after HpaII digestion indicating that the repeat expansion is not methylated (Figure 2.1a). To confirm that MspJI is able to cut the hexanucleotide repeat if it is methylated, DNA from a repeat expansion carrier was *in vitro* methylated using CpG methyltransferase (M. SssI) and digested with MspJI. Repeat primed-PCR analysis showed poor amplification of *in vitro* methylated DNA relative to unmethylated DNA (Fig 2.2a), demonstrating the ability of MspJI to cut methylated GGGGCC repeats.

We then extended our methylation analysis to include the two CpG islands adjacent to the repeat expansion (open boxes in Figure 2.1b). A bisulfite cloning screen was performed on cerebellar DNA from four *C9orf72* expansion carriers and four nonexpanded controls to identify

methylated CpG dinucleotides. PCR was used to amplify three regions covering both CpG islands from bisulfite converted DNA (amplicons A-C, Figure 2.1b). Amplified DNA was cloned, and individual clones (≥five per case) were sequenced. Increased methylation (10-25% of clones) was observed in the vicinity of the 5' CpG island upstream of the hexanucleotide repeat region (amplicon A, Figure 2.1b). The second CpG island downstream of the repeat expansion was not methylated in both repeat expanded and control cases (amplicons B and C, Figure 2.1b). The sequencing results for all clones are provided in Fig 2.2b-d. These results confirm recently published studies demonstrating promoter hypermethylation in some repeat expansion mutation cases (Xi et al., 2013).

Since methylation was observed only in *C9orf72* mutation cases, we hypothesized that promoter methylation occurs in cis to the repeat expansion. To determine whether methylation was monoallelic vs. biallelic, we identified three hypermethylated *C9orf72* mutation cases heterozygous for the rs200034037 polymorphism, a dinucleotide deletion with an allele frequency of ~25.6%. This polymorphism lies upstream of HhaI and HpaII restriction enzyme cut sites within the differentially methylated region (DMR). Both HhaI and HpaII only cut unmethylated DNA. The DMR was amplified after HhaI/HpaII double digestion and subject to PAGE electrophoresis (Figure 2.1c). In all three cases, only the major rs200034037 allele was amplified after digestion of unmethylated DNA, in contrast with mock digested DNA in which both major and minor alleles were observed. These results were confirmed by Sanger sequencing which showed biallelic sequences after the rs200034037 polymorphism in mock digested DNA. This is in contrast with HhaI/HpaII digested DNA which showed only the major allele sequence (color traces, bottom of Figure 2.1c). Thus, *C9orf72* promoter methylation is monoallelic, consistent with promoter hypermethylation in cis relative to the repeat expansion.

Methylation of Repeat Expansion Carrier Brains

To confirm the presence of promoter methylation in a larger cohort of subjects, we developed an alternative method based on the HhaI restriction enzyme recognition site within the

DMR (Figure 2.3a). HhaI digestion of unmethylated DNA was coupled with quantitative PCR (qPCR) which demonstrated a shift in DNA amplification curves relative to mock digested DNA (Figure 2.3b). This shift was used to calculate the percent DNA that is resistant to HhaI digestion as a measure of methylation status. This method proved to be robust, as standard curves using *in vitro* methylated DNA showed high linearity from 0 to 100% methylation (Figure 2.3c).

This assay was used to confirm the presence of *C9orf72* promoter methylation using DNA from the frontal cortex and cerebellum of repeat expanded and nonexpanded cases. Both the frontal cortex and cerebellum of *C9orf72* mutation carriers exhibited higher methylation than controls (p=0.0009, Figure 2.3d). Despite this highly significant result, *C9orf72* promoter methylation for many *C9orf72* mutation carriers was low and within the normal range, suggesting that *C9orf72* promoter methylation is not driving disease pathogenesis.

C9orf72 RNA Expression in Lymphoblast Cell Lines

We analyzed ENCODE CAGE-seq data to determine whether there are differences in *C9orf72* transcription across diverse cell lineages (Rosenbloom et al., 2013). *C9orf72* expression is highly variable across different cell lines, with both hematopoietic and neuronal cell lines expressing similar levels of *C9orf72* in contrast with epithelial, fibroblastic or other embryonic or mesenchymal cell types (Figure 2.4a). CAGE-seq data also indicated that of the three annotated *C9orf72* mRNAs, variant 2 (V2) represented the major RNA transcript, representing 92.6 +/- 1.2% of all *C9orf72* transcripts (Figure 2.4b). Based on these analyses, we turned to lymphoblast cell lines (LCLs) to demonstrate that methylation is associated with reduced *C9orf72* expression. LCLs from non-expanded (ND16183) versus expanded ALS cases (ND11836, ND10966 and ND14442) were analyzed by Southern blotting using a probe specific to *C9orf72* which confirmed the presence of large repeat expansions in the three mutant LCLs (Figure 2.4c). Methylation was very low (<1%) in non-expanded ND16183 cells and was high in two expanded cell lines (ND10966 and ND14442, Figure 2.4d). Despite harboring the hexanucleotide repeat expansion,

ND11836 cells were not methylated, akin to our results above which showed many carriers exhibit minimal *C9orf72* promoter methylation.

C9orf72 mRNA was measured using RT-qPCR which demonstrated that the unmethylated but repeat expanded cell line, ND11836, showed essentially normal total *C9orf72* mRNA levels (Figure 2.4e). Despite expressing normal levels of total *C9orf72* mRNA, ND11836 cells demonstrated increased usage of the alternative upstream TSS, as shown by the increase in V3 *C9orf72* mRNA relative to control cells (Figure 2.4g). Since transcription from the upstream TSS results in transcription through intron 1 which contains the hexanucleotide repeat, we speculated that the repeat expansion mutation is associated with accumulation of intron 1-containing RNA. Indeed, using primers 5' to the hexanucleotide repeat within intron 1, unmethylated expanded cells exhibited over 4-fold more intronic *C9orf72* RNA relative to control cells (Figure 2.4h).

In contrast, expanded cells with *C9orf72* promoter hypermethylation exhibited reduced total (Figure 2.4e) and V2 (Figure 2.4f) mRNA relative to control cells. Promoter methylation was also associated with reduced expression from the alternative TSS, as determined by reduced V3 mRNA expression relative to unmethylated expanded cells (Figure 2.4g). The reduction of V3 mRNA expression was coupled to reduced intronic RNA accumulation in methylated expanded cells relative to unmethylated expanded cells (Figure 2.4h).

To verify that the *C9orf72* mutation causes a shift in TSS usage and to confirm that methylation inhibits *C9orf72* RNA levels, the effect of DNA demethylation was tested in hypermethylated expanded LCLs using the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (5-aza-dC). Repeat expanded lymphoblast cells (ND14442) and nonexpanded control cells (ND16183) were treated with 5-aza-dC for 3 days followed by a 3 day recovery without 5-aza-dC. 5-aza-dC treatment resulted in a 41.8% reduction in *C9orf72* promoter methylation ($p<0.0001$, Figure 2.5a). 5-aza-dC treatment increased total ($p<0.05$, Figure 2.5b) and V3 ($p<0.001$, Figure 2.5d) *C9orf72* mRNA expression in expanded but not control cells, demonstrating that *C9orf72* promoter methylation is linked to transcriptional silencing of mutant

C9orf72. However, V2 mRNA was not significantly changed in expanded cells upon demethylation (Figure 2.5c). Thus, promoter demethylation reverses transcriptional silencing of *C9orf72* but not of the major V2 transcript. Rather, demethylation rescued only total and V3 mRNA indicating that the expanded allele is preferentially transcribed from the upstream TSS. This alternative TSS usage translated into a marked increase in intronic *C9orf72* accumulation ($p<0.01$, Figure 2.5e). Thus methylation inhibits the accumulation of potentially toxic intronic RNA by inhibiting transcription through the repeat expansion.

Selective Vulnerability of Hypomethylated Cells

The accumulation of repeat-containing intronic RNA has been postulated to promote neurodegeneration. Since promoter hypermethylation inhibited intronic RNA levels, we hypothesized that hypermethylation may protect cells from cellular insults, such as oxidative stressors which lead to stress granule formation or altered proteostasis. Both processes have been implicated in ALS and FTD (Y. R. Li et al., 2013; S. C. Ling et al., 2013). To test this hypothesis, expanded and control cells were demethylated with 5-aza-dC and exposed to sodium arsenite or chloroquine, known to induce stress granules or inhibit autophagy, respectively. Toxicity was assessed by measuring lactate dehydrogenase (LDH) release into media. In both control and expanded cell lines, 5-aza-dC resulted in mild toxicity compared to untreated cells consistent with the known toxicity profile of 5-aza-dC (control $p<0.01$, mutant $p<0.05$, Figure 2.5f-g). Importantly, demethylation of control non-expanded cells with 5-aza-dC had no effect on arsenite or chloroquine toxicity, indicating that global hypomethylation does not result in an altered response to these cellular stressors (Figure 2.5f). In contrast, hypomethylation of mutant cells with 5-aza-dC resulted in enhanced sensitivity to both arsenite and chloroquine (arsenite $p<0.05$, chloroquine $p<0.001$, Figure 2.5g). Thus reactivation of mutant *C9orf72* transcription which results in the accumulation of mutant intronic *C9orf72* RNA appears to be associated with increased vulnerability to cellular stressors.

Hypermethylation, RNA foci and RANT Pathology

Having demonstrated that *C9orf72* promoter hypermethylation was associated with lower levels of intronic RNA using RT-qPCR, we hypothesized that hypermethylation is associated with reduced accumulation of RNA foci. LCLs were studied using in situ hybridization with a fluorescently-labeled locked nucleic acid (LNA) probe that recognizes the GGGGCC hexanucleotide repeat. RNA foci were observed in expanded cell lines but not in non-expanded cells (Figure 2.6a). LCLs were scored for the presence of RNA foci which demonstrated that 12% of ND11836 cells (expanded and unmethylated) exhibited RNA foci, with several cells exhibiting multiple RNA foci (Figure 2.6c). In contrast, expanded and methylated cell lines, ND10966 and ND14442, exhibited significantly fewer RNA foci compared to ND11836 (1.5% for ND10966, $p<0.0001$; 3% for ND14442, $p=0.00094$, Fisher's exact test, Figure 2.6c).

Human brain tissue was also used to test whether *C9orf72* hypermethylation is associated with reduced accumulation of RNA foci. Intact nuclei were isolated from post-mortem cerebellum and stained for RNA foci. RNA foci were only observed in repeat expanded cases and not in non-expanded control cases (Figure 2.6b). The percentage of nuclei containing RNA foci were quantified from 12 repeat expanded cases which revealed an inverse relationship between *C9orf72* methylation and the percentage of nuclei with RNA foci ($p = 0.0148$, Figure 2.6d).

RANT aggregates accumulate as a consequence of non-ATG translation of mutant *C9orf72* RNA which may contribute to disease pathogenesis (Ash et al., 2013; Mori et al., 2013). Given that hypermethylation reduced mutant *C9orf72* expression, it seemed likely that hypermethylation would be associated with reduced RANT pathology. To understand the relationship between methylation and RANT pathology, cerebellar tissue sections from hypomethylated (<12% HhaI resistance) and hypermethylated (>12% HhaI resistance) repeat expansion carriers were stained with antibodies specific for GA (glycine-alanine), GP (glycine-proline), or GR (glycine-arginine) repeats corresponding to the three different sense reading frames of the repeat expansion. The 12% threshold corresponds to three standard deviations

above the mean HhaI resistance values for control, non-expanded cases. Immunohistochemistry showed abundant numbers of inclusions in hypomethylated *C9orf72* expansion cases and markedly fewer RANT inclusions in hypermethylated cases (Figure 2.7a). Quantification of the percent of cerebellar granule neurons with RANT inclusions demonstrated a significant reduction in RANT pathology burden associated with *C9orf72* promoter hypermethylation (Figure 2.7b). To confirm this finding, insoluble proteins were extracted from hypermethylated and hypomethylated cerebellum and assayed by dot blot analysis which again demonstrated that hypermethylation was associated with reduced accumulation of GA, GP, and GR proteins (Figure 2.7c). These results show a tight relationship between *C9orf72* promoter methylation and RANT pathology, and indicate that endogenous transcriptional silencing reduces the downstream pathologic features associated with the hexanucleotide repeat expansion.

Discussion:

We examined the downstream consequences of *C9orf72* promoter hypermethylation and provide evidence that *C9orf72* promoter methylation found in a subset of *C9orf72* expansion carriers may be protective against toxicity associated with the *C9orf72* hexanucleotide repeat expansion. Hypermethylation (Xi et al., 2013) and repressive histone marks (Belzil, Bauer, et al., 2013) have been previously observed in the promoter region of *C9orf72* in repeat expansion carriers, but the downstream consequences of these changes were not known. We found that (1) *C9orf72* promoter but not the hexanucleotide repeat is hypermethylated in a subset of mutation carriers in cis with the repeat expansion mutation, (2) the repeat expansion is associated with alternative TSS usage, (3) promoter methylation leads to transcriptional gene silencing of both V2 and V3 *C9orf72* transcripts, (4) methylation inhibits the accumulation of potentially toxic intronic RNA in LCLs and brain, (5) demethylation of *C9orf72* promoter appears to be associated with increased vulnerability to cellular stressors, and (6) hypermethylation is associated with reduced accumulation of RNA foci and RANT aggregates.

Repeat expansion mutations can cause disease through a variety of means including gain and loss of function mechanisms. Mounting evidence suggests that the *C9orf72* mutation results in a toxic gain of function linked to the accumulation of either mutant intronic RNA within RNA foci, or the accumulation of RANT aggregates. In this regard, the *C9orf72* mutation may be similar to the trinucleotide repeat expansion mutation affecting *DM1* associated with myotonic dystrophy where repeat expanded RNA accumulate as RNA which sequesters muscleblind protein (Miller et al., 2000b). In fact, Pur α (Z. Xu et al., 2013), hnRNPA3 (E. B. Lee et al., 2013; Z. Xu et al., 2013), hnRNP-H and ADARB2 (Stepto, Gallo, Shaw, & Hirth, 2014) have been shown to bind to GGGGCC repeats and may be sequestered within RNA foci, resulting in altered RNA metabolism. Furthermore, toxicity through increased vulnerability to cellular stressors caused by autophagy inhibitors and glutamate in induced pluripotent stem cell-derived neurons are linked to expression of the hexanucleotide repeat (Almeida et al., 2013; Donnelly et al., 2013). Hexanucleotide repeat length also correlates with toxicity in neuroblastoma cells (Y.-B. Lee et al.,

2013). Collectively, these data support the hypothesis that the accumulation of mutant intronic RNA through transcription of the repeat expansion contributes to vulnerability to cellular stress.

Alternatively, the repeat expansion mutation has also been associated with a reduction in *C9orf72* mRNA and protein together with various epigenetic marks indicative of transcriptional silencing (Belzil, Bauer, et al., 2013; Waite et al., 2014a; Xi et al., 2013). This would be analogous to Fragile X syndrome where repeat expansion mutations in *FMR1* cause dense promoter hypermethylation and transcriptional silencing, resulting in a loss of function (O'Donnell & Warren, 2002; Sutcliffe et al., 1992). Notably, both nonsense and missense *FMR1* mutations are also associated with Fragile X syndrome, supporting the loss of function hypothesis (Collins, Bray, et al., 2010; Collins, Coffee, et al., 2010; Gronskov, Brondum-Nielsen, Dedic, & Hjalgrim, 2011; Myrick et al., 2014). In contrast, nonsense and missense *C9orf72* mutations were not found in a large ALS cohort (M. B. Harms et al., 2013). Moreover, individuals homozygous for the *C9orf72* mutation are phenotypically similar to heterozygous carriers, which argues against a loss of function mechanism (Cooper-Knock et al., 2013; P. Fratta et al., 2013). Experimental data is conflicting regarding *C9orf72* haploinsufficiency, as reduction of *C9orf72* in mice with antisense oligonucleotides does not result in a neurodegenerative phenotype in mice while knockdown of the *C9orf72* homologue in zebrafish results in motor neuron axonal degeneration (Ciura et al., 2013b; Lagier-Tourenne et al., 2013).

We demonstrate here that these two mechanisms may not be independent of each other, but rather that transcriptional silencing of *C9orf72* appears to inhibit the formation of RNA foci and RANT aggregates (Figure 2.8). Given that promoter methylation can suppress hexanucleotide repeat expansion associated pathologies, we propose that epigenetic silencing of *C9orf72* may be a protective counter-regulatory response to the presence of the hexanucleotide repeat expansion. This is in contrast to epigenetic silencing seen in Fragile X syndrome, where silencing of *FMR1* causes disease. While additional studies are clearly needed to determine the functional consequences of reduced *C9orf72* expression, our observation that demethylation of the *C9orf72*

promoter appears to be associated with reduced cellular vulnerability supports the hypothesis that the *C9orf72* mutation causes disease due to a deleterious gain of function.

Our study identifies two molecular mechanisms which contribute to the accumulation of repeat-expanded intronic RNA. First, the presence of the repeat expansion mutation is associated with increased usage of the upstream TSS. Expression from this alternative TSS results in expression of the hexanucleotide repeat. We observed alternative TSS usage in unmethylated mutant LCLs, and confirmed this finding using 5-aza-dC treated LCLs which showed that demethylation of the *C9orf72* promoter reactivates V3 but not V2 expression. These results are consistent with a recent report showing alternative TSS usage in repeat-expanded neurons differentiated from induced pluripotent stem cells (Sareen et al., 2013). Alternative usage of the upstream TSS results in transcription through the hexanucleotide repeat expansion within intron 1, resulting in increased intronic RNA accumulation and the formation of RNA foci. The reason for alternative TSS usage is not clear, although the hexanucleotide repeat is known to form very stable configurations including G quadruplexes that may alter local chromatin structure and affect transcription (Pietro Fratta et al., 2012; Haeusler et al., 2014; Reddy, Zamiri, Stanley, Macgregor, & Pearson, 2013). The second mechanism that regulates intronic RNA accumulation is promoter methylation. Indeed, promoter methylation was associated with reduced levels of *C9orf72* transcripts regardless of which TSS was used. Reduced transcription led to reduced intronic RNA accumulation and reduced numbers of RNA foci. This supports a model in which the repeat expansion mutation leads to downstream molecular pathologies which are inhibited by promoter methylation in a subset of cells and individuals (Figure 2.8).

RANT aggregates have been proposed to be toxic (Ash et al., 2013; Mori et al., 2013), perhaps through disruption of cellular pathways like autophagy or altered proteostasis (Cleary & Ranum, 2013; Pearson, Nichol Edamura, & Cleary, 2005). However, RANT pathology does not associate with severity of neurodegeneration and it is unclear whether RANT aggregates are pathogenic (Mackenzie et al., 2013). Regardless, we observed a strong correlation between hypermethylation and reduced *C9orf72* RANT pathology within hypermethylated *C9orf72*

mutation carriers, again indicating that promoter hypermethylation is an endogenous pathway which inhibits hexanucleotide repeat expansion associated aberrations. Indeed, *C9orf72* promoter hypermethylation is the first known epigenetic modification that appears to inhibits pathology in a neurodegenerative disease.

If methylation indeed modulates *C9orf72* mutation pathways, *C9orf72* methylation status may be used as a diagnostic biomarker to separate individuals into epigenetic subtypes. *C9orf72* mutation carriers are clinically heterogeneous, ranging from rapidly fatal motor neuron disease to the so-called "slowly progressive" form of frontotemporal dementia (Boeve et al., 2012; Cooper-Knock, Hewitt, et al., 2012; Hsiung et al., 2012; Khan et al., 2012; Murray et al., 2013; Simon-Sanchez et al., 2012). The molecular basis for this heterogeneity is largely unknown. Hexanucleotide repeat length measured from peripheral blood does not appear to correlate with disease phenotypes (Benussi et al., 2013; Dols-Icardo et al., 2013; van Blitterswijk et al., 2013). Due to the relatively low numbers of cases studied here, we did not observe any clinical parameters associated with *C9orf72* hypermethylation.

In other polynucleotide expansion diseases, hypermethylation in the vicinity of repeat expansions have been linked to clinical outcomes. In Fragile X Syndrome, *FMR1* methylation inversely correlates with intelligence quotient scores (de Vries et al., 1996; R. J. Hagerman et al., 1994; Loesch et al., 1993; McConkie-Rosell et al., 1993; Merenstein et al., 1996; Reiss, Freund, Baumgardner, Abrams, & Denckla, 1995; Smeets et al., 1995). Similarly, *FXN* methylation in Freidreich's ataxia also inversely correlates with disease severity (Castaldo et al., 2008; Evans-Galea et al., 2012b). In both Fragile X syndrome and Freidreich's ataxia, repeat expansion mutations result in a loss of function mutation that cause disease, consistent with hypermethylation leading to a more severe disease phenotype. In the case of *FMR1*, the repeat expansion is thought to result in epigenetic silencing of the mutant allele in cis, linked to the formation of RNA-DNA hybrids in early development (Colak et al., 2014; Devys et al., 1992; Eiges et al., 2007). Whether this represents a general mechanism responsible for hypermethylation adjacent to repeat expansion loci remains to be determined. In contrast with Fragile X syndrome

and Freidreich's ataxia, our study supports the hypothesis that the *C9orf72* mutation causes disease through a gain of function mutation. Therefore, we predict that *C9orf72* hypermethylation may be associated with milder clinical disease. It will be important to determine whether *C9orf72* methylation correlates with clinical parameters relevant to disease in a larger cohort of *C9orf72* mutation carriers.

Experimental models have also demonstrated heterogeneity in terms of *C9orf72* RNA expression and molecular phenotypes. Given that methylation affects many of the downstream effects of the hexanucleotide repeat mutation, current efforts to develop and characterize cellular or other experimental models should include promoter methylation analysis (Almeida et al., 2013; Donnelly et al., 2013; Sareen et al., 2013). Methylation status may be a confounding factor when trying to compare cell lines derived from different patients. Indeed, we observed significant differences between LCLs based on methylation status. Reported discrepancies regarding *C9orf72* expression between various cell lines may similarly be linked to differences in methylation (Sareen et al., 2013).

Finally, determining whether *C9orf72* hypermethylation indeed modifies disease progression in *C9orf72* mutation carriers will help guide the development of novel molecular therapies. Given that the *C9orf72* mutation reduces gene expression, one approach may be to increase *C9orf72* expression to reverse any deleterious effects of low *C9orf72* expression (Ciura et al., 2013b). While it remains possible that reduced expression of *C9orf72* is deleterious, our results suggest that efforts to increase *C9orf72* expression as a potential therapy for hexanucleotide repeat expansion carriers should proceed with caution as such approaches may lead to increased accumulation of potentially toxic RNA. In contrast, our results bolster recent efforts to develop molecular therapies for *C9orf72* mutation carriers based on antisense oligonucleotides that target hexanucleotide repeat expanded RNA for post-transcriptional degradation (Donnelly et al., 2013; Lagier-Tourenne et al., 2013; Sareen et al., 2013). Another therapeutic possibility includes development of therapies that promote transcriptional silencing of mutant *C9orf72,* thereby reducing toxic RNA and downstream RANT pathology.

In summary, a subset of *C9orf72* mutation carriers demonstrate *C9orf72* promoter hypermethylation, which may represent an endogenous protective response to the hexanucleotide repeat expansion. Promoter hypermethylation results in stable silencing of the mutant gene and reduction in the downstream pathologies associated with the *C9orf72* mutation. Since transcriptional silencing is associated with a protective phenotype, this study supports the hypothesis that the *C9orf72* hexanucleotide repeat expansion causes disease by a gain of toxic function as opposed to haploinsufficiency, and highlights an endogenous molecular pathway which may be amenable to future therapy development.

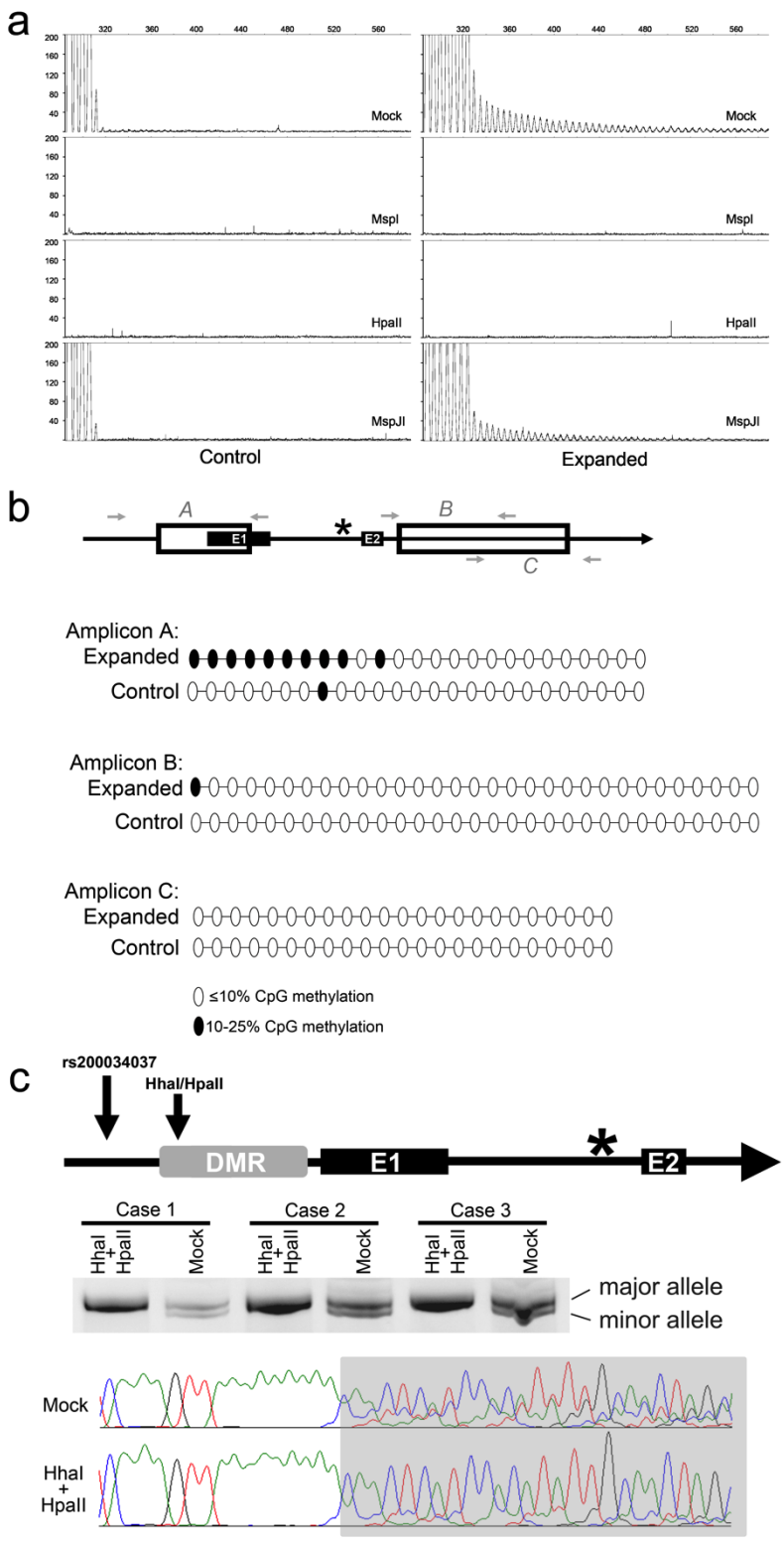
a
Mock
MspI
HpaII
MspJI
Control
Expanded
b
A
B
C
E1
E2
Amplicon A:
Expanded
Control
Amplicon B:
Expanded
Control
Amplicon C:
Expanded
Control
≤10% CpG methylation
10-25% CpG methylation
c
rs200034037
HhaI/HpaII
DMR
E1
E2
Case 1
Case 2
Case 3
HhaI+HpaII
Mock
major allele
minor allele
Mock
HhaI + HpaII

Fig. 2.1: Hypermethylation of the *C9orf72* Promoter
(a) Cerebellar DNA from control (n=8, left) and repeat expanded cases (n=8, right) were mock digested (no enzyme) or digested with MspI, HpaII or MspJI. DNA was subject to repeat primed PCR and representative electropherograms are shown.
(b) Top panel shows a schematic of the bisulfite sequenced regions where filled boxes are exons, open boxes are CpG islands, and the star is the GGGGCC repeat expansion. Amplicon A covers the first CpG island, amplicon B covers the first half of the second CpG island and amplicon C covers the second half of the second CpG island. The bottom panels are summaries of bisulfite cloning results in which cerebellar DNA from 4 *C9orf72* repeat expansion carriers and 4 control cases (n=20-21 clones per genotype) were sequenced. Each oval represents a single CpG dinucleotide where unfilled oval represents an unmethylated CpG dinucleotide (0-10% of clones) and a filled oval represents a methylated CpG dinucleotide (10-25% of clones). Methylation over 25% was not observed.
(c) Top panel shows a schematic of the 5' end of *C9orf72* including the differentially methylated region (DMR, shaded) upstream of the 1st coding exon (E1) of *C9orf72.* The dinucleotide deletion polymorphism (rs200034037) and HhaI/HpaII cut sites are shown as arrows and the star is the hexanucleotide repeat expansion upstream of the 2nd coding exon (E2). DNA from *C9orf72* promoter hypermethylated repeat expanded cases (n=3) that contain the polymorphism were mock digested (no enzyme) or digested with HpaII and HhaI. DNA from case 1 is from a lymphoblastoid cell line (ND14442) while DNA from cases 2 and 3 are from peripheral blood. The region flanking the deletion and restriction enzyme cut sites were amplified and run on a polyacrylamide gel to separate the major vs. minor rs200034037 alleles. Representative sequencing chromatograms of mock digested or HhaI/HpaII digested DNA are shown where the gray area denotes the sequences demonstrating monoallelic vs. biallelic sequences downstream of rs200034037.

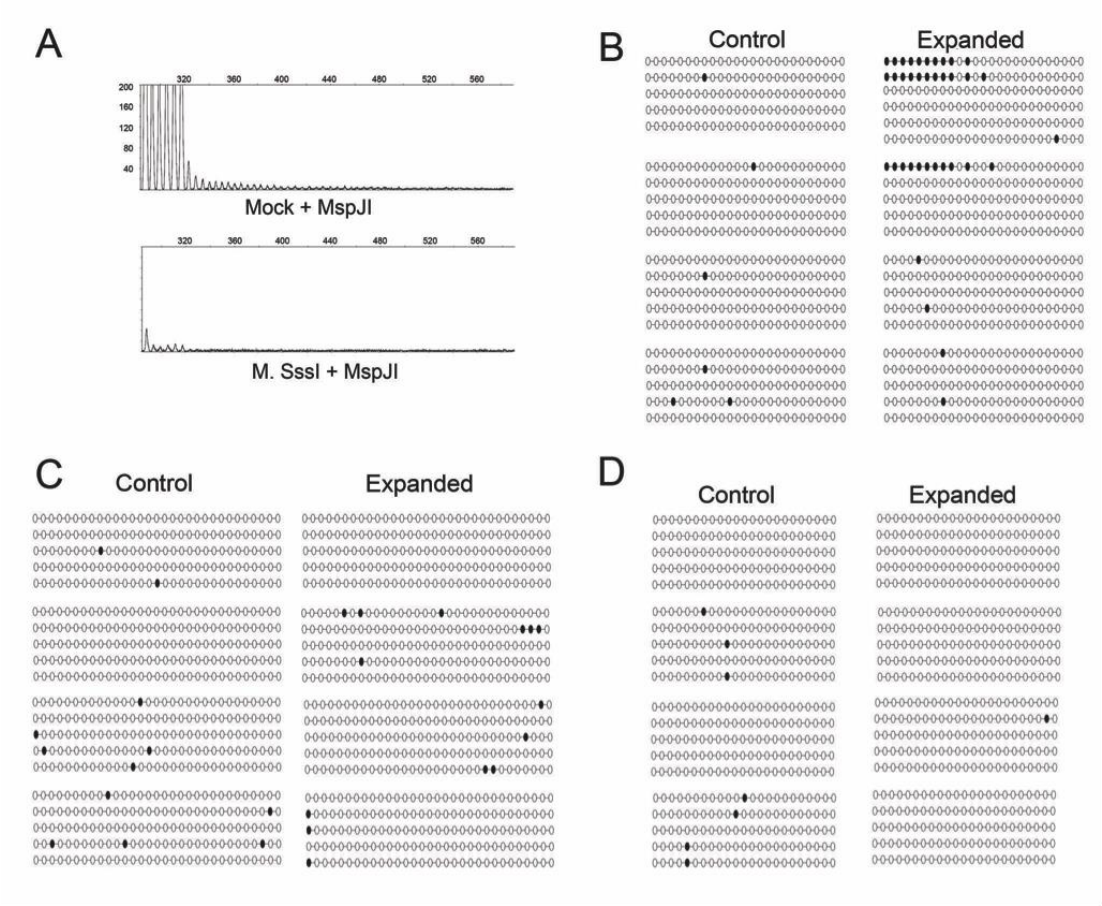

Figure 2.2: Repeat primed PCR and bisulfite cloning of *C9orf72*
(a) DNA from a repeat expanded case was in vitro methylated using MSssI methyltransferase and digested with MspJI. Digested material was subject to repeat primed PCR and representative electropherograms are shown.
(b) Bisulfite cloning results for each sequenced clone for amplicon A (panel B), amplicon B (panel C) and amplicon C (panel D) are shown. Four non-expanded controls and four expanded cases were assessed. Each oval represents a single CpG dinucleotide where unfilled ovals represent unmethylated CpG dinucleotides and filled ovals represent methylated CpG dinculeotides. Sequencing results are presented in groups of 5-6 clones which correspond to results from each case. Background non-CpG cytosines showed a conversion rate of >95%.

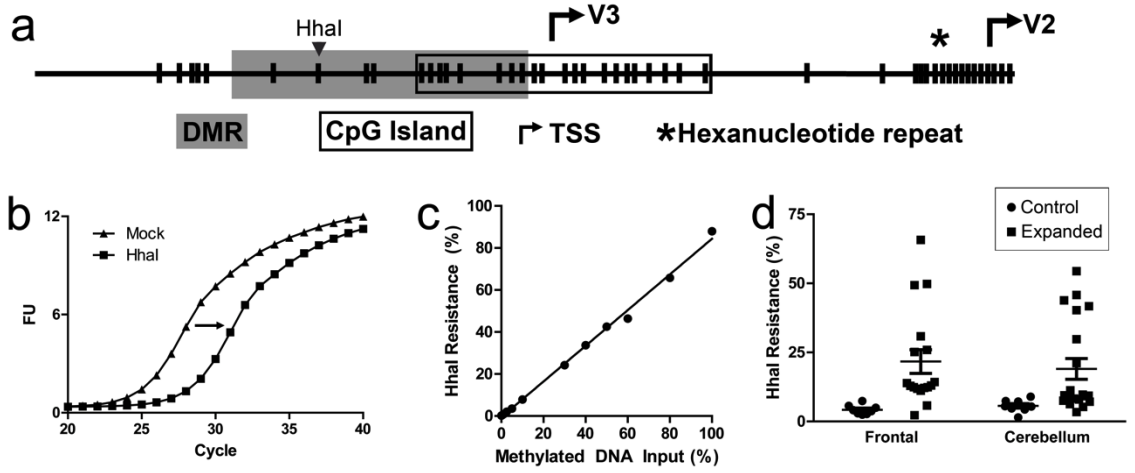

Fig. 2.3: *C9orf72* Promoter Hypermethylation in Repeat Expanded and Control Brain
(a) Schematic representation of the 5' end of the *C9orf72* gene in which individual CpG dinucleotides are designated by vertical bars, The upstream CpG island is designated with an open box, the TSS for V2 and V3 transcripts are designated by arrows, and the hexanucleotide repeat region is designated by a star. The differentially methylated region (DMR) is shaded. The HhaI restriction enzyme recognition site in the differentially methylated region is shown.
(b) Representative qPCR amplification curves for mock (no enzyme) versus HhaI digested DNA demonstrating a shift in the amplification curve upon DNA digestion. The magnitude of this shift is used to calculate the % DNA resistant to HhaI digestion as a measure of DNA methylation.
(c) DNA from control LCLs was *in vitro* methylated with MSssI, and various ratios of methylated and mock methylated DNA were tested. Digest qPCR quantification for HhaI resistance was plotted for increasing amounts of *in vitro* methylated DNA input.
(d) HhaI digest resistance as assessed by digest-qPCR of frontal cortex (n=8-17) or cerebellum (n=8-20) DNA from control (circles) or repeat expansion cases (squares) are shown. Individual values are plotted in addition to the mean and standard error. Two-way ANOVA: genotype p=0.0009, region p=0.8851, interaction p=0.6445.

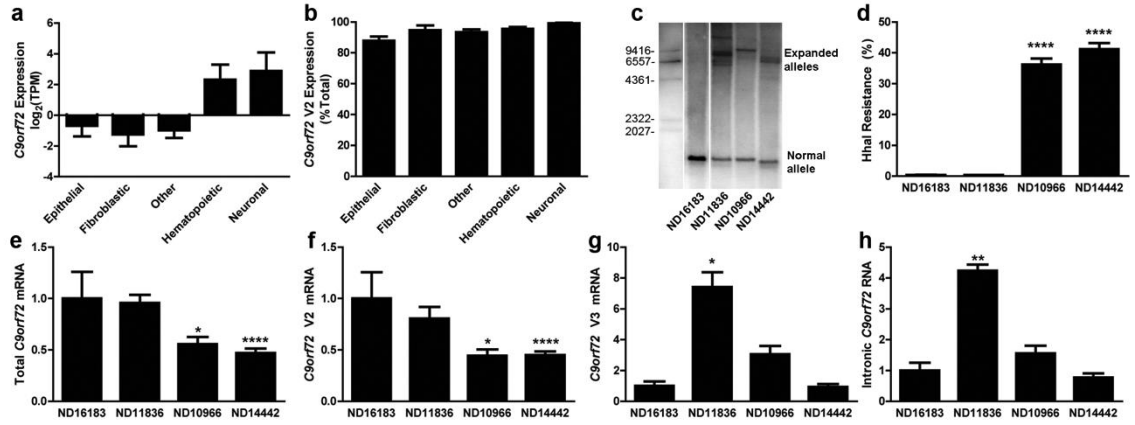

Fig. 2.4: *C9orf72* Methylation Inhibits Expression of Mutant RNA
(a) ENCODE CAGE-seq quantification of *C9orf72* mRNA. The total number of *C9orf72* sequence tags were normalized for number of total sequence reads, shown as mean log2 transformed tags per million (TPM) +/- SE. Cell lines were divided into different cell lineages as labeled, with other representing various mesenchymal and embryonic stem cell lineages.

(b) ENCODE CAGE-seq quantification of *C9orf72* V2 mRNA relative to total *C9orf72* mRNA expression. The number of V2 tag sequences were normalized to total *C9orf72* tag sequences, shown as mean % of total +/- SE.
(c) Southern blot of LCL DNA from non-expanded (ND16183) and expanded (ND11836, ND10966 and ND14442) cultures using a probe specific for *C9orf72* that recognizes the normal allele (bottom) and expanded alleles (top). Molecular weight markers are shown as indicated.
(d) HhaI resistance from non-expanded (ND16183) and expanded (ND11836, ND10966 and ND14442) LCLs, shown as mean + SE. One-way ANOVA: $p<0.0001$. ****$p<0.0001$ relative to ND16183. Each cell line was measured in triplicate.
(e-h) RT-qPCR quantification shown as mean + SE (n=5) for total mRNA (e), V2 mRNA (f), V3 mRNA (g) and intronic RNA (h) from control (left) or expanded (right) lymphoblast cells. *$p<0.05$, ** $p<0.01$, **** $p<0.0001$.

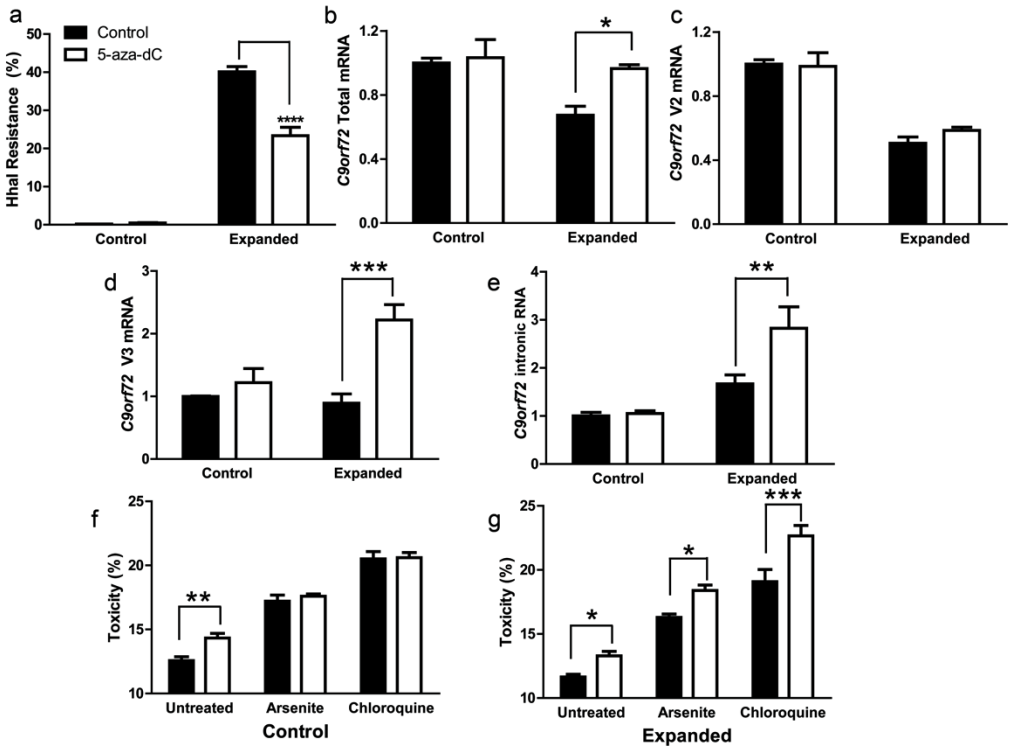

Fig. 2.5: *C9orf72* Promoter Demethylation Promotes Toxic RNA Accumulation
(a) HhaI resistance shown as mean + SE (n=5) from control (left) or expanded (right) lymphoblast cells treated with 5-aza-dC (open) or untreated (filled). Two-way ANOVA: genotype p<0.0001, treatment p<0.0001, interaction p<0.0001. ****p<0.0001
(b-e) RT-qPCR quantification shown as mean + SE (n=5) for total mRNA (b), V2 mRNA (c), V3 mRNA (d) and intronic RNA (e) from control (left) or expanded (right) lymphoblast cells treated with 5-aza-dC (open) or untreated (filled). Two-way ANOVA for total (genotype p=0.0087, treatment p=0.0264, interaction p=0.0681), V2 (genotype p<0.0001, treatment p=0.5197, interaction p=0.3579), V3 (genotype p=0.026, treatment p=0.0006, interaction p=0.0076) and intronic RNA (genotype p=0.0001, treatment p=0.0259, interaction p=0.0397). *p<0.05, **p<0.01
(f-g) Control (f) and expanded cells (g) that were untreated (filled) or 5-aza-dC treated (open) were assessed for toxicity (%LDH release relative to untreated, mean + SE, n=3-6) after 24 hours of no additional treatment (left), arsenite (10 µM, middle) or chloroquine (100 µM, right). Two-way ANOVA for control cells (5-aza-dC p=0.0358, stressor p<0.0001, interaction p=0.0773) and mutant cells (5-aza-dC p<0.0001, stressor p<0.0001, interaction p=0.1471). *p<0.05, **p<0.01, ***p<0.001

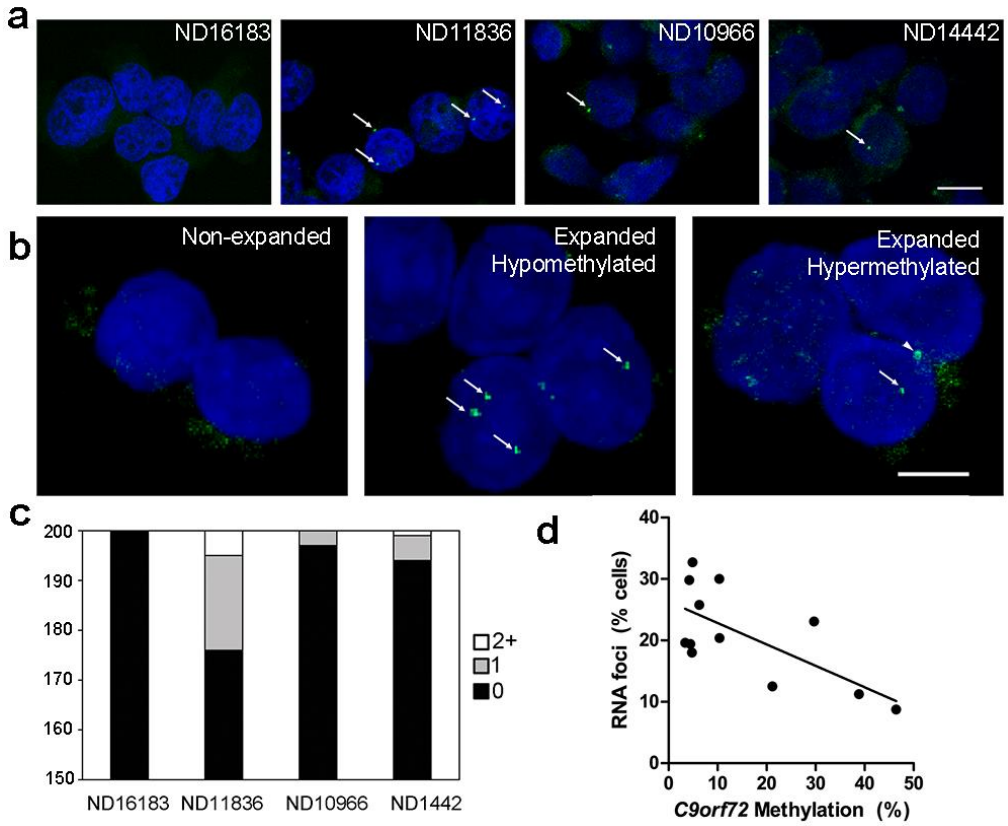

Fig. 2.6: *C9orf72* promoter hypermethylation inhibits RNA foci accumulation
(a) Representative *in situ* hybridization images of non-expanded (ND16183), unmethylated and expanded (ND11836), and hypermethylated and expanded (ND10966 and ND14442) LCLs. RNA foci in green are highlighted with arrows, and nuclei are counterstained blue with DAPI. Scale bar = 10 µm.
(b) Representative *in situ* hybridization images of non-expanded (left), hypomethylated and expanded (middle), and hypermethylated and expanded (right) cerebellar nuclei. RNA foci in green are highlighted with arrows, and nuclei are counterstained blue with DAPI. Arrowhead points to non-specific autofluorescence. Scale bar = 5 µm.
(c) The number of RNA foci in LCLs was scored (n=200 cells per cell line), shown as a stacked bar graph to demonstrate the proportion of LCLs with zero, one or multiple RNA foci.
(d) The percent of cerebellar nuclei with RNA foci from 12 repeat expanded cases is shown as a function of *C9orf72* methylation. A linear regression was performed ($R^2 = 0.4633$, $p = 0.0148$).

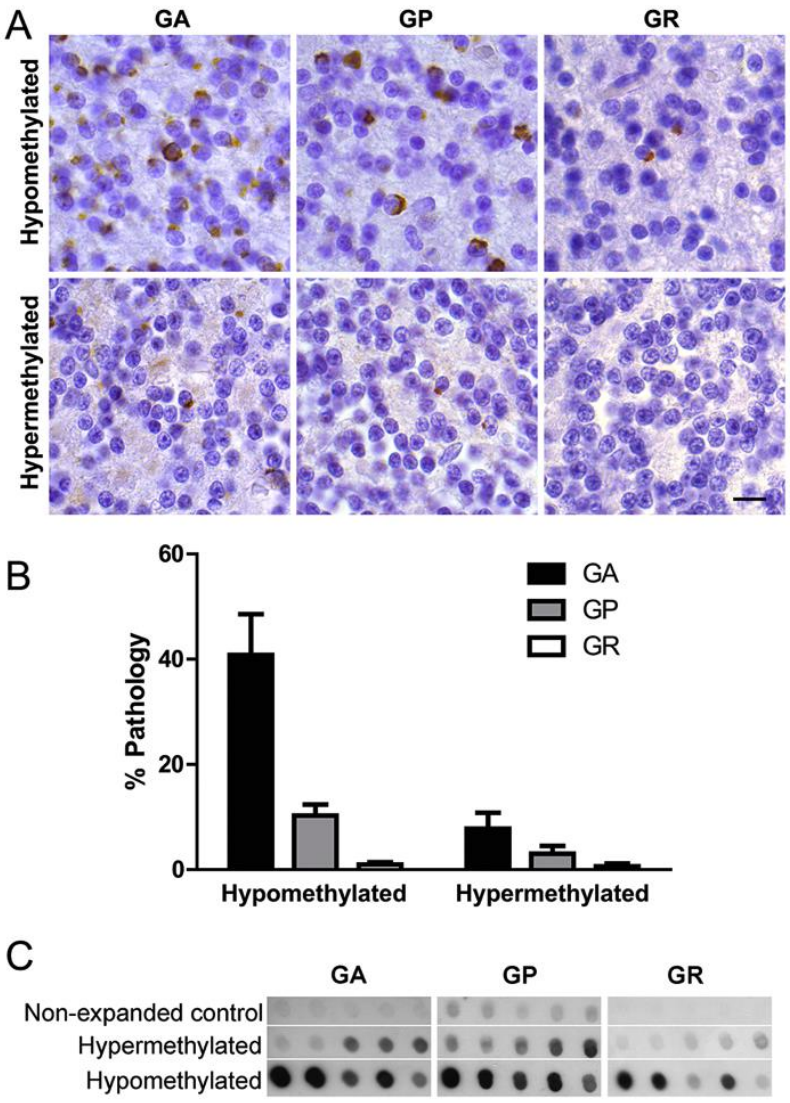

Fig. 2.7: *C9orf72* promoter hypermethylation and RANT accumulation in *C9orf72* repeat expansion carriers.
(a) Representative immunohistochemistry of cerebellar granular neurons using antibodies recognizing glycine-alanine (GA), glycine-proline (GP) and glycine-arginine (GR) dipeptide repeat proteins. Aggregates appear brown while nuclei are counterstained with hematoxylin. Scale bar = 10 µm.
(b) Quantification of RANT pathology in 11 hypomethylated and 7 hypermethylated *C9orf72* repeat expansion carriers shown as mean + standard error of the percentage of cerebellar granular neurons containing GA (left), GP (middle) or GR (right) aggregates. Two-way ANOVA: $p<0.0001$ (pathology type), $p=0.0080$ (methylation), 0.0002 (interaction).
(c) Sarkosyl-insoluble material was biochemically extracted and subject to dot blot analysis using antibodies that recognize GA, GP or GR dipeptide repeat proteins in 5 nonexpanded controls (top row), 5 hypermethylated *C9orf72* repeat expansion carriers (middle row), and 5 hypomethylated *C9orf72* repeat expansion carriers (bottom row).

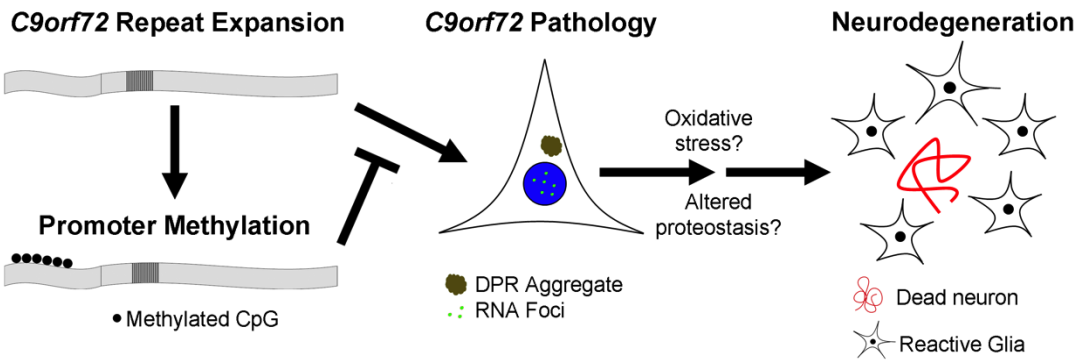

Fig. 2.8: *C9orf72* Promoter Methylation Model
C9orf72 repeat expansions are associated with promoter hypermethylation in subset of mutation carriers. Expression of the repeat expansion leads to the accumulation of mutation-specific pathologies, namely RNA foci and DPR aggregates. The accumulation of mutant RNA is associated with increased vulnerability to cellular stressors, including oxidative and autophagic stress. Promoter hypermethylation inhibits these downstream effects by reducing the accumulation of RNA foci and/or DPR aggregates.

Acknowledgements

Cell lines (ND16183, ND11836, ND10966, and ND14442) and clinical data from the NINDS Repository (ccr.coriell.org/ninds) were used. We thank Dr. Linda Kwong, Yan Xu and the Center for Neurodegenerative Disease Research for providing RANT antibodies and autopsy materials. The authors would like to thank the patients and patients' families who made this research possible.

Funding Sources

This study was supported in part by a grant from the Judith & Jean Pape Adams Foundation and by the National Institutes of Health (K08AG039510, T32AG00255, P30AG10125, P01AG017586, P01AG032953).

CHAPTER 3:

Loss of Nuclear TDP-43 Is Associated with Reactivation of LINE elements

Abstract:

Loss of nuclear TAR DNA binding protein 43 (TDP-43) into cytoplasmic aggregates is a characteristic pathology observed in amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) post-mortem human brain. TDP-43 is a ubiquitous nuclear RNA binding protein that functions broadly in RNA processing. Because TDP-43 pathology is the strongest correlate to neurodegeneration, it is important to understand the effects of nuclear TDP-43 loss and how it leads to disease. Using a novel subcellular fractionation method and fluorescent activated cell sorting to enrich for diseased neuronal nuclei without TDP-43 and normal neuronal nuclei with TDP-43 from post-mortem ALS/FTD human brain, I was able to characterize the effects of TDP-43 loss on the transcriptome and chromatin accessibility. I show that loss of nuclear TDP-43 contributes to (1) global gene expression changes linked to pathways including RNA processing, (2) dysregulation of introns and 3' untranslated regions, and (3) massive splicing alterations. I was able to show altered TDP-43 activity via altered *TARDBP* autoregulation and selective vulnerability of superficial neocortical neurons. Here, I have validated previous experimental discoveries related to TDP-43 biology using human brain. Given that TDP-43 has been implicated in histone dysregulation and chromatin dynamics, chromatin accessibility in pathologic TDP-43 nuclei was investigated. Indeed, in post-mortem ALS/FTD human brain, nuclear TDP-43 loss is associated with altered chromatin accessibility around repetitive long interspersed nuclear elements (LINE) and also increased LINE1 DNA. Furthermore, loss of TDP-43 leads to increased retrotransposition, suggesting that loss of TDP-43 contributes to toxicity via reactivation of transposable elements.

Introduction:

Amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) are neurodegenerative diseases with overlapping clinical, genetic and pathologic features. ALS is a debilitating neuromuscular disease that is characterized by the loss of motor neurons in the motor cortex and spinal cord, whereas FTD is a fatal neurocognitive disease that preferentially affects the frontal and temporal lobes. Importantly, a significant subset of ALS patients can develop FTD and vice versa (Strong et al., 2017). Many genetic mutations have been linked to both diseases (reviewed in S. C. Ling et al., 2013) but even in genetically homogenous populations, the mechanisms of neurodegeneration are not entirely clear. A key similarity and pathological hallmark of both diseases is the loss of nuclear TDP-43 into cytoplasmic aggregates (Neumann et al., 2006).

TDP-43 is a nuclear RNA binding protein that shuttles between the nucleus and cytoplasm and has been shown to be involved in multiple facets of RNA processing, ranging from mRNA splicing, transcription, transport, and stability (Lee et al., 2012). TDP-43 binds to non-coding RNAs including introns and 3' UTRs (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011; Tollervey et al., 2011) and repetitive elements including transposable elements and long interspersed nuclear elements (LINE) (W. Li, Jin, Prazak, Hammell, & Dubnau, 2012). LINE elements make up at least 17% of the human genome (Lander et al., 2001) and can be active in the human genome within somatic cells (Muotri et al., 2005). While typically silenced, LINE elements may be reactivated in response to different cellular stresses, including aging or in the setting of a few neurologic diseases (Bundo et al., 2014; Coufal et al., 2011; Muotri et al., 2010). Increased endogenous retroviral activity has also been observed in ALS patients (Bowen et al., 2016; McCormick et al., 2008; Steele et al., 2005), raising the possibility that aberrant TDP-43 activity can influence the expression of genomic repeat elements that are typically silenced.

Both the loss of nuclear TDP-43 and the gain of cytoplasmic TDP-43 aggregates are presumed to be neurotoxic. However, as TDP-43 pathology is the strongest correlate to

neurodegeneration, efforts have been made in animal and cellular models to study how both mechanisms can contribute to neurodegeneration (Huang et al., 2012; Igaz et al., 2011; Scekic-Zahirovic et al., 2016; A. Sharma et al., 2016; Tsai et al., 2010; Wils et al., 2010; Y. F. Xu et al., 2010). Unfortunately, these models have been unable to fully recapitulate both facets of TDP-43 pathology, leading to the use of human pathologic tissue to infer disease mechanisms (R. H. Batra, K.; Vu, A.;Rabin, S.J.; Baughn, M.W.; Libby, R.T.; Hoon, S.; Ravits, J.; Yeo, G.W., 2016; Cooper-Knock, Kirby, et al., 2012; Prudencio et al., 2015). Importantly, the full spectrum of molecular changes associated with TDP-43 pathology in human neurons is not known. While post-mortem brain from ALS/FTD patients can be used to study disease pathogenesis, molecular analysis of post-mortem human brain is associated with several challenges. Considering the human neocortex is a heterogeneous mix of cell types with an approximately 75% glial population, whole neocortex molecular analyses of neurodegenerative disease tissues often reflects average changes from reactive glial cells rather than diseased neurons (Azevedo et al., 2009; Srinivasan et al., 2016). Indeed, transcriptome analysis of ALS/FTD brain tissue identified changes in inflammatory processes, which are likely secondary to directly pathogenic disease mechanisms (Prudencio et al., 2015).

Fluorescence activated cell sorting (FACS) can be used to isolate rare events for molecular analysis. Herein, we developed a novel fractionation method based on subcellular fractionation and FACS isolation of diseased neuronal nuclei without TDP-43 versus uninvolved normal neuronal nuclei from post-mortem human ALS/FTD brains to identify the role of nuclear TDP-43 loss on the transcriptome and chromatin accessibility. We applied this method to show that predicted features of TDP-43 loss are actually occurring in diseased neurons including global gene expression changes involving pathways like RNA processing, dysregulation of introns and 3' UTR segments, abundant splicing changes, altered *TARDBP* autoregulation, and selective vulnerability of superficial neocortical neurons. Furthermore, we identified increased chromatin accessibility around LINE elements and increased LINE1 DNA in neuronal nuclei without TDP-43 from post-mortem brain, and demonstrate that the loss of TDP-43 protein is associated with

increased LINE-1 retrotransposition. These results suggest that loss of TDP-43 contributes to toxicity via reactivation of transposable elements.

Materials and Methods:

Clinical, genetic, and pathological assessments.

Human autopsy tissue was obtained from the University of Pennsylvania Center for Neurodegenerative Disease Research Neurodegenerative Disease Brain Bank as described (Toledo et al., 2014). Clinical and demographic information is available as Table 3.1.

Human brain nuclei isolation and FACS.

Mid-frontal neocortex was dounce homogenized using pestil B (Kimble Chase, Rockwood, TN, USA) in 0.25M sucrose in TKM (50mM Tris, 25mM KCl, 5 mM MgCl2) buffer. The homogenate was adjusted to 1.6M using 2.3M sucrose in TKM. The homogenate was spun on a 1.8M sucrose cushion in TKM at using a SW41 rotor on the Beckman Coulter XPN-80 ultracentrifuge at 40,000g for 40 minutes at 4^{o}C (Beckman Coulter Inc, Indianapolis, IN, USA). Isolated nuclei were stained with Alexa Fluor 647 conjugated to 2089 **(**rabbit polyclonal C-terminal anti-TDP-43 antibody, Center for Neurodegenerative Disease Research, University of Pennsylvania), Alexa Fluor 488 conjugated NeuN (EMD Millipore, Billerica, MA, USA), and DAPI (Invitrogen, Carlsbad, CA, USA)**.** Alexa Fluor 647 was conjugated to 2089 according to the APEX Alexa Fluor 647 Antibody labeling kit protocol (Thermo Fisher Scientific, Waltham, MA, USA). Stained nuclei were sorted for single cells based on DAPI, NeuN and TDP-43 fluorescence on the BD FACSAria II (BD Biosciences, San Jose, CA, USA) at 20 psi on 100µm nozzle.

RNA Isolation and RNA-seq Library Generation.

Nuclei were sorted into Buffer RLT and RNA was extracted using the standard AllPrep DNA/RNA Micro kit protocol (Qiagen, Germantown, MD, USA). RNA was amplified using the Ovation V2 RNA-Seq System V2 (NuGEN, San Carlos, CA, USA) according to the manufacturer's protocol and resulting cDNA was quantified by Nanodrop 2000 (Thermo Fisher). cDNA was sheared using the Covaris S2 (Covaris, model S2, Woburn, MA, USA) to 200 base

pairs and the resulting cDNA library was created using the Ovation Ultralow Library System kit (NuGEN) according to the manufacturer's protocol. The size distribution of the library was assessed using BioAnalyzer Lab Chip 1000 (Agilent, Santa Clara, CA, USA), and quantified using the Qubit dsDNA kit (Invitrogen) and Kapa Biosystems Library Quantification kit (KapaBiosystems, Boston, MA). cDNA libraries were pooled, clustered on the cBot and sequenced using 100 or 125 base pairs paired end reads on the HiSeq 2000 or 2500 (Illumina, San Diego, CA, USA).

Pre-processing, mapping and filtering of RNA-seq data.

Raw sequencing reads were demultiplexed through the University of Pennsylvania Functional Genomics Core. FastQC (Version 0.11.3; (Andrews, 2010)) was used to assess sequencing read quality. Reads were mapped to the human genome (GRCh38, GENCODE release 22; (Harrow et al., 2012)) using STAR (Version 2.2.4 with option --outFilterIntronMotifs RemoveNonCanonical; (Dobin et al., 2013)) with default parameters and only uniquely mapping reads were selected for further analysis. Ribosomal and mitochondrial reads were removed using modified scripts from the PORT pipeline (https://github.com/itmat/Normalization). Coverage plots for all uniquely mapped reads were computed using sam2cov (https://github.com/khayer/sam2cov/). SAM files were converted to BAM files using samtools view and BAM files were sorted by coordinate with samtools sort.

***Creation of non-overlapping gene, exon and intron annotations*.**

Annotations were based on the comprehensive gene annotation file of the GENCODE Release 22 (GRCh38.p2). The GTF file downloaded from GENCODE (http://www.gencodegenes.org/releases/22.html") was loaded into R (Version 3.2.2; (Team, 2015), R Foundation for Statistical Computing, Vienna, Austria) and converted into a TranscriptDb object with the makeTranscriptDbFromGFF tool in the Bioconductor package GenomicFeatures (Version 1.22.8; (Lawrence et al., 2013)). From the TranscriptDb object, all

annotated Ensembl genes and their exons were obtained using exonsBy (by="gene"). Ensembl gene ids were replaced by official gene symbols using biomaRt (Version 2.26.1). Genes with several Ensembl gene IDs were combined into one record. For each duplicated gene, overlapping exons were combined into single exons. Thus, genes were defined as the sequence between the first base of the first exon and the last base of the last exon. Introns were defined as the gap regions between non-overlapping exons. Furthermore, regions shared by overlapping genes were removed to count reads that map to one gene or to genic elements (exon or intron) from one gene. Exons and introns from the same gene were numbered sequentially and used for subsequent analyses in the R package DEXSeq (1.16.10).

Differential gene and genic element expression analysis.

The mapped, filtered RNA sequencing reads were counted using a custom R script including the R packages Rsamtools (1.22.0), GenomicFeatures (1.22.8) and GenomicAlignments (1.6.3; (Lawrence et al., 2013)). Briefly, sorted BAM files were loaded into R (3.2.2) using readGAlignmentPairs and the number of reads mapping to genes, mapping to exons or introns, was computed using findOverlaps (with options type="within" and ignore.strand=TRUE) and countSubjectHits. Genes were analyzed for differential expression using the R package DESeq2 (1.10.1; (Love, Huber, & Anders, 2014)). Paired analysis was used to determine differences between TDP-positive (TDPpos) and TDP-negative (TDPneg) nuclei (option design=~experiment+condition in DESeqDataSetFromMatrix tool where experiment is the subject id). Batch effects, mainly caused by gender, were identified with the R package sva (3.18.0) and removed according to the DESeq2 vignette.

Differential genic element expression analysis was done using DEXSeq (1.16.10). Briefly, the sorted BAM files were used to count the number of reads mapping to exons and introns using the exon and intron annotation generated above and computed using findOverlaps (with options type="within" and ignore.strand=TRUE) and countSubjectHits. Reads mapping to exon-intron junctions were excluded. We controlled for gender as potential confounding factor by including it

in the linear model used in the analysis by adding the term “gender:exon” and performed a paired analysis. The following linear models were used full model = ~sample + exon + gender:exon + experiment:exon + condition:exon and reduced model = ~sample + exon + gender:exon + experiment:exon and reduced model = ~sample + exon + gender:exon, where sample is the sample id and experiment the subject id. In addition, dispersions were estimated using the tool estimateDispersions with option fitType=’local’. Changes in expression were significant if Bonferroni-Hochberg multiple testing adjusted p-values were less than 0.05.

WGCNA and Gene Ontology Analysis.

Gene ontology analysis of significantly differentially expressed genes with gene symbols was performed using EnrichR according to 2015 gene ontology terms (Kuleshov et al., 2016). Pathways with an adjusted p-value < 0.05 were considered significantly enriched pathways. Weighted gene co-expression network analysis was done using the R package DESeq2 and WGCNA (1.51). First, counts were normalized and transformed using functions estimateSizeFactors, DESeq and getVarianceStabilizedData. The pickSoftThreshold function was used for the selection of the soft thresholding power β. A power of 16 was chosen which was the lowest power for which the scale-free topology fit index reaches 0.9. This was then used to derive the adjacency matrix from the normalized counts. To minimize effects of noise and spurious associations, the adjacency matrix was transformed into a Topological Overlap Matrix using the function TOM and the corresponding dissimilarity (1-TOM) was calculated. Next hierarchical clustering and the branch cutting algorithm Dynamic Tree Cut (function cutreeDynamic with minimum cluster size of 21) to identify co-expression modules. Modules with high correlation (>0.8) were merged into a common module using function mergeCloseModules. The resulting module eigengenes of each module were correlated with TDP-43 status to identify modules whose co-expression can be linked to TDP-43 status using the function cor (with option use=”p”) and corPvalueStudent (to calculate p-values). Genes from one module that was significantly correlated with TDP-43 status were identified based on whether they were enriched for pathways

involving ubiquitin processing, RNA splicing and DNA repair as identified by EnrichR. The similarity matrix and connections for these genes were then output into VisANT using the function exportNetworkToVisANT. The nodes are color coded as described in the legend.

Alternative splicing analysis.

All FASTQ files were trimmed to 100bp using a Python script, aligned to GRCh38 using STAR, and ribosomal and mitochondrial reads were removed. The resulting SAM files were converted to sorted BAM files using samtools and rMATS.3.0.9 (Shen et al., 2014) was run using default parameters with the following options (-t paired –len 100 –c 0.05 – analysis P). Significant alternative splicing events were used if Bonferroni-Hochberg multiple testing adjusted p-values were less than 0.05.

RNA binding protein CLIP analysis.

Published TDP-43 iCLIP data from SH-SY5Y cells (Tollervey et al., 2011) and hnRNP A, A2B1, F, M, U CLIP data from HEK293 cells (Huelga et al., 2012) was used to analyze TDP-43 or hnRNP binding sites. Raw hg18 Bowtie files were converted to SAM files using samtools with the bowtie2sam.pl and then converted to FASTQ files in UNIX. The resulting files were aligned to hg38 transcriptome using STAR with the following parameters (--runThreadN 8 --outSAMAttribute All --outFilterMultimapScoreRange 0 --alignIntronMax 50000 --outFilterMatchNMin 15 --outFilterMatchNminOverLread 0.9 --outSAMtype BAM SortedByCoordinate). PIPECLIP (Chen B et al. Gen Biol. 2014) was run using the python script with the sorted BAM file with the following parameters (-l 25 –m 2 –c 3 –r 0 –M 0.01 –c 0.01 –s hg38). The resulting BED files were used for subsequent analysis. Using R, the “GenomicRanges” package was used to determine whether coordinates of RNA binding protein sites overlapped with differentially used bin genomic coordinates using the “findOverlaps” function. The number of genomic bins with RNA binding protein sites were quantified and used to test for enrichment of differentially used bins with RNA

binding protein sites with chi-square tests performed on GraphPad Prism (GraphPad, San Diego, CA, USA). Significant enrichment was detected with p-values less than 0.05.

Generation of ATAC-seq libraries.

ATAC-seq libraries were generated using an established protocol (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013) with slight modifications. Briefly, 30,000 sorted nuclei were centrifuged at 1,000g for 10 min at 4°C. Pellets were resuspended in Tn5 transposase reaction mix (Illumina) and incubated at 37°C for 30 min followed by purification using MinElute PCR purification kit (Qiagen). Following purification, library fragments were amplified using the Nextera index kit (Illumina), using the following cycling conditions: 72°C for 5 min, 98°C for 30 sec, followed by 5 cycles of 98°C for 10 sec, 63°C for 30 sec and 72°C for 1 min. Amplification optimization was performed using a small aliquot of the reaction mixture. In total, libraries were amplified between 11-16 cycles. Next libraries were size selected for fragments in size of 150 to 1200 bp using SPRIselect beads (Beckman-Coulter). The size distribution of the libraries was assessed using the BioAnalyzer Lab Chip 1000 and quantified by Qubit HS DNA kit as well as by qPCR (KAPA Biosystems). Libraries were pooled and sequenced on the HiSeq 2500 (Illumina) generating 100 bp single-end reads.

Preprocessing, mapping and filtering of ATAC-seq data.

Sequencing reads were demultiplexed by the University of Pennsylvania Functional Genomics Core and the raw read FASTQ files from four sequencing runs were merged for each sample using the bash command cat. Sequencing read quality was assessed with FastQC (Version 0.11.3). Nextera paired-end adapter sequences were removed from reads using Trimmomatic in single end mode (Version 0.36 with option -phred33, seedMismatches=2, palindromeClipThreshold=30, simpleClipThreshold=10 and option MINLEN:36; Bolger, Lohse & Usadel 2014). Only reads with a minimum length of 36 bp were kept after trimming and were aligned against the human genome (GRCh38, GENCODE release 25) using STAR (Version 2.2.4

with option --alignIntronMax 1 and --alignMatesGapMax 1800; Dobin et al. 2013) with default parameters. Reads that (1) mapped to more than one locus; (2) mapped to the mitochondrial genome; and (3) were read duplicates were excluded. Filtering was done with samtools (0.1.19) and MarkDuplicates (option REMOVE_DUPLICATES=true) from the Picard tools (2.3.0) was run for the deduplication. Next, the resulting bam files were converted to bed files using bedtools (2.24.0). Finally, reads were offset by +4 bp if they mapped to the positive strand and by -5 bp if they mapped to the negative strand. Thereby, the 5'-end of the end matches the Tn5 transposase cut site.

Identification and annotation of differentially accessible genomic regions.

The human autosomes were divided into 1,000 bp long non-overlapping bins using bedtools makewindows tool. Next, reads mapping to each bin were counted using findOverlaps and countSubjectHits functions from the R package GenomicAlignments. Bins with less than 4 average read count over all 14 samples were excluded from the analysis. Next the Bioconductor package edgeR (3.12.1) was used to identify differentially accessible genomic regions. Briefly, differences in library composition due to differences in sequencing depth were adjusted for by calculation of a scaling factor with the function calcNormFactors from the edgeR package. Paired analysis (option design=~patient+condition where patient is the subject id and option robust=TRUE in function estimateDisp) and a generalized linear model (GLM) likelihood ratio test was used to determine differences between TDPpos and TDPneg nuclei (glmFit function).

Annotation of bins with genomic and repeat elements.

For genomic partition analysis, UCSC knownGene annotation GFF files and RepeatMasker repeat annotation files for the hg38 genome build were downloaded from the UCSC table browser. A custom awk script was then used to extract 5' UTR exons and introns, 3' UTR exons and introns, and exons and introns for each protein coding gene. Promoter annotations were defined as 1,000bp upstream of the first exon in the transcript, either coding or

in the UTR. The GFF files were then converted to BED format and parsed to only include canonical chromosomes (chr1-22, chrX, chrY), after which all overlapping exons were merged together (bedtools v2.18.1 merge command) and the intervals were sorted. Then, mutually exclusive genomic element annotations were generated using a hierarchical scheme where any base pair overlapping multiple annotations is assigned to a class following the hierarchy 5' UTR exon > 5' UTR intron > 3' UTR exon > 3' UTR intron > promoter > mRNA exon > mRNA intron > repeat, and if it did not fall into any of these categories it was classified as intergenic. This was implemented with a custom bash script using awk and the bedtools complement and intersect commands. Each set of differentially accessible genomic bins was then compared against these annotations using the bedtools intersect command. First, bins were annotated against the mutually exclusive annotations with a requirement that the annotation overlapped at least 51% of the bin (-f 0.51). Any bin not overlapping a 5' UTR exon or intron, 3' UTR exon or intron, promoter, mRNA exon, or mRNA intron was then re-overlapped against the original RepeatMasker annotations with a 16% overlap requirement to define the full set of bins overlapping repeat elements. Any repeats overlapping bins with this criteria were annotated with RepeatMasker repeat family annotations when available.

LINE1 qPCR assay.

LINE1 DNA content was determined with an established LINE1 qPCR assay protocol (Coufal et al., 2009) with slight modifications. Briefly, nuclei were extracted from frozen post-mortem brain of seven FTD/ALS patients (the same patients that were used for RNA-seq and ATAC-seq) and FACS-sorted according to TDP-43 status. For each patient, an equal number of TDP-positive and TDP-negative neuronal nuclei was used for genomic DNA extraction. 70,000-105,000 nuclei were sorted into lysis buffer RLY from the NucleoSpin DNA RapidLyse kit (Macherey-Nagel, Bethlehem, PA, USA). According to the manufacturer's instructions, proteinase K was added and nuclei were incubated at 56°C for 1 hour in a thermomixer set to 800 rpm. Based on the total volume of the proteinase K digest reaction the volume of binding buffer RLB

was increased. Each sample was spun through the provided spin columns. Washing steps were done according to the manufacturer's instructions. Genomic DNA was eluted with buffer RLE and genomic DNA concentrations were determined with the Qubit high sensitivity dsDNA kit. Oligonucleotide PCR primers were purchased from Sigma Genosys (Sigma-Aldrich, St Louis, MA, USA) and TaqMan-MGB probes from Applied Biosystems (Thermo Fisher Scientific). Primers against L1ORF2 (Forward: 5' TGCGGAGAAATAGGAACACTTTT 3' and Reverse: 5' TGAGGAATCGCCACACTGACT 3') and SATA (Forward: 5' GGTCAATGGCAGAAAAGGAAAT 3' and Reverse: 5' CGCAGTTTGTGGGAATGATTC 3') described in (Coufal et al., 2009) were used. Taqman probes for L1ORF2 (VIC-CTGTAAACTAGTTCAACCATT-MGBNFQ) and SATA (6-FAM-TCTTCGTTTCAAAACTAG-MGBNFQ) were also used in the same master mix reaction. qPCR experiments were performed on the ABI StepOne Plus (Thermo Fisher Scientific) and using TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific). The multiplexing reaction was optimized by limiting reaction components: 150 nM of the L1ORF2 primers, 900 nM of the SATA primers and 250 nM of the TaqMan probes were used in each qPCR reaction. Each qPCR reaction was run using five different gDNA concentrations (72 pg, 76 pg, 80 pg, 84 pg and 88 pg) and in triplicate. We required qPCR reactions to exhibit minimal to no variance in cycles to threshold (Ct) values for L1ORF2 and SATA between technical replicates, and minimal to no differences in Ct values for SATA between matched TDPpos and TDPneg nuclei. These criteria were chosen as TDPpos vs. TDPneg nuclei were extracted from the same patient material and therefore should have minimal variation in the number SATA repeats. LINE1 DNA content was determined by obtaining the ratio of ΔCt values for L1ORF2 over SATA and compared between TDP-positive and TDP-negative nuclei by using a paired t-test.

LINE1 Retrotransposition Assay.

The retrotransposition assay protocol was performed as described with minor modifications (Kopera et al., 2016). HeLa cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum, 100U/ml penicillin-streptomycin (Pen/Strep), 2mM

L-glutamine at 37°C and 5% CO_2. Briefly, HeLa cells were seeded at density of 150,000 cells per 6 well overnight. Using Fugene 6 (Promega, Madison, WI), cells were either co-transfected with pEF06R (Addgene plasmid #42940, (Farkash, Kao, Horman, & Prak, 2006)) and pcDNA5/TO, wildtype TDP-43, or TDP-43 with 2 mutated nuclear localization signals 24 hours after seeding cells to determine whether TDP-43 overexpression can affect LINE1 retrotransposition. To test the effects of TDP-43 loss, cells were transfected with a plasmid expressing the Cas9 enzyme (Addgene plasmid # 62988, (Ran et al., 2013)) and 2 different guide RNAs targeting the *TARDBP* genes (TDP-43 KO1 top: 5' CACCGCCCATGGAAAACAACCGAAC 3'; TDP-43 KO1 bottom: 5' AAACGTTCGGTTGTTTTCCATGGGC 3'; TDP-43 KO2 top: 5' CACCGACATCCGATTTAATAGTGTT 3'; TDP-43 KO2 bottom: 5' AAACAACACTATTAAATCGGATGTC 3'). Concurrently, control cells were co-transfected using the same plasmids except that pMSCV-PIG (Addgene plasmid #21654, (Mayr & Bartel, 2009) was used instead of pEF06R as a GFP control. 24 hours after co-transfection, the transfection reaction was stopped by exchanging the medium. Puromycin selection at 1.5µg/mL was started 3 days after transfection and replaced on day 5, followed by puromycin selection at 0.75µg/mL on day 7. Cells were collected on day 9 after transfection for flow cytometry. Cells were trypsinized, washed, and resuspended in DMEM without phenol red and 2% FBS with propidium iodide (PI) solution for live/dead analysis. Stained cells were analyzed using the BD FACSCanto machine (BD Biosciences). Cells were gated for single cells using side scatter vs forward scatter, then for PI and GFP fluorescence using FlowJo (FlowJo, Ashland, OR, USA). GFP positive cells were normalized to all viable (PI negative) cells to calculate the percentage of GFP positive cells. Triplicate cultures were assessed over three to four independent experiments. A mixed-effects linear regression model using the R package "nlme" was used to determine whether there was a significant relationship between TDP-43 expression and retrotransposon activity where the experiment was the random variable. Relationships with a p value < 0.05 were considered significant.

Genetic Background	Sex	Age of Death	Age of Onset	PMI	Diagnosis	MND	Dementia
C9orf72	F	71	62	3.5	FTLD	N	Y
C9orf72	M	57	55	8.5	FTLD-ALS	Y	Y
C9orf72	F	61	57	12	FTLD	N	Y
C9orf72	F	73	N/A	N/A	FTLD	N	Y
C9orf72	M	77	71	5	FTLD	N	Y
C9orf72	M	57	55	6	FTLD-ALS	Y	Y
C9orf72	M	75	71	10	FTLD-ALS	Y	Y

Table 3.1: Clinical demographics of patient cohort.

Results:

Fluorescence Activated Cell Sorting of Pathologic Neuronal Nuclei

To determine the molecular alterations associated with TDP-43 pathology, FACS was used to sort for neuronal nuclei from the mid-frontal cortex of post-mortem ALS/FTD patient brains with TDP-43 pathology (Fig 3.1A). The frontal cortex was used due to the abundant TDP-43 pathology found in this brain region (Neumann et al., 2006). Patient demographics of this patient cohort can be found in Table 1. Nuclei were isolated from archived frozen brain tissue and stained for NeuN to identify neurons and TDP-43 to identify pathologic cells. Flow cytometry plots of non-diseased control individuals show two populations: (1) NeuN negative, TDP-43 positive nuclei (non-neuronal); and (2) NeuN positive, TDP-43 positive nuclei (neuronal). This is in contrast to the three distinct populations from ALS/FTD patients: (1) NeuN negative, TDP-43 positive nuclei (non-neuronal); (2) NeuN positive, TDP-43 positive nuclei (unaffected neuronal); and (3) NeuN positive, TDP-43 negative nuclei (pathologic neuronal, Fig 3.1B). On average, 7.05% of all neuronal nuclei (<2% of all cells) were TDP-43 negative nuclei.

Validation for FAC sorting was done using both (1) confocal imaging of sorted neuronal nuclei with and without TDP-43 and (2) differential gene expression between sorted and unsorted nuclei to determine enrichment of neuronal genes. Imaging of sorted nuclei showed that TDP-43 positive and negative nuclei both express NeuN but either have nuclear TDP-43 or have lost nuclear TDP-43, respectively (Fig 3.1C). Furthermore, sequencing of RNA from an equal number of unsorted vs. sorted TDP-43 positive neuronal nuclei was done from two human cases to show that sorting enriched for neurons as indicated by an enrichment of neuronal and depletion of non-neuronal genes from sorted nuclei. Coverage plots for a gene expressed in astrocytes and oligodendrocytes respectively including glial fibrillary acidic protein (*GFAP*) and myelin basic protein (*MBP*) showed a marked depletion of reads in neuronal nuclei compared to unsorted nuclei. In contrast, sorted nuclei had more reads mapped to neuronal genes like glutamate decarboxylase 2 (*GAD2*) (Fig 3.1D). To determine whether there was global enrichment of neuronal genes within sorted nuclei, differential gene expression analysis was done using

DESeq2 to compare gene expression between sorted and unsorted nuclei. Among 1071 differentially expressed genes, 575 genes were annotated as neuronal or non-neuronal genes based on previous annotations of neuronal and non-neuronal genes (Y. Zhang et al., 2014). Out of 22 significantly upregulated genes in sorted nuclei, there was a significant enrichment of neuronal genes ($\chi 2$ odds ratio=14.7958; p-val=1.221e-05). There was a corresponding depletion of non-neuronal genes within sorted nuclei (1039 out of 1049, $\chi 2$ odds ratio=13.40478; p-val<2.2E-16) (Fig 3.1E). Thus, FACS sorting from human post-mortem brain was able to deplete glial transcripts while enriching for neuronal populations.

RNA-Sequencing of Pathologic Neuronal Nuclei

Seven cases of ALS/FTD were subjected to FAC sorting to obtain RNA from equal numbers of TDP-43 positive and negative nuclei (35,000 to 100,000 nuclei per sample). To ensure genetic homogeneity, only patients bearing the *C9orf72* hexanucleotide repeat expansion were used. The clinical demographics of the patients are found in Table 3.1. RNA was used to prepare libraries for 100 or 125bp paired end sequencing on the Illumina HiSeq 2000/2500. Between 14 libraries, there was a total of 1.971 billion raw reads and after mapping to Gencode GRCh38, a total of 1.126 billion reads were uniquely mapped with a mean of 80,400,486 uniquely mapped reads per library. On average, 60.84% of reads mapped to introns, 11.53% of reads mapped to exons and 27.63% of reads mapped to intergenic regions. This was consistent with the fact that sequencing nuclear RNA yields more intronic than exonic reads (Ameur et al., 2011; Mitchell et al., 2012).

Loss of TDP-43 in human brain results in dysregulated RNA expression.

Because TDP-43 is a nuclear RNA binding protein and serves multiple functions within the cell, we hypothesized that TDP-43 loss would lead to RNA dysregulation. Differential gene expression analysis using DESeq2 of neuronal nuclei with TDP-43 (TDPpos) and without TDP-43 (TDPneg) was done. Principal component analysis was used to assess data variability and to

ensure that changes within the dataset were driven by the presence or absence of TDP-43. PCA of the entire dataset showed that principal component 1 (PC1) explained 21% of the variation whereas PC2 explained 19% of the variation (Fig 3.2A). Interestingly, when each patient was paired (shown as lines that connect pairs of TDPpos and TDPneg nuclei from the same individual), the slope of each case was similar, suggesting that changes associated with TDP-43 pathology was consistent between individual patients. Thus, the variation explained by PC1 and PC2 is in large part associated with the loss of TDP-43.

Differential gene expression analysis using DESeq2 showed that out of 31,898 expressed genes, 5576 were significantly differentially expressed genes (DEG) due to the loss of TDP-43 (2849 upregulated and 2727 downregulated genes, Fig 3.2B). Furthermore, significantly differentially expressed genes have significantly larger gene (t-test, $p<2.2E-16$) and intron sizes (t-test, $p<2.2E-16$), consistent with the fact that TDP-43 binds to long pre-mRNAs (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011) (Fig 3.2C-D).

Gene ontology analysis using EnrichR of significantly upregulated genes in TDPneg nuclei demonstrated that there was an enrichment of genes related to transport mechanisms, synaptic transmission, mRNA processing, protein kinase and histone regulation (Table 3.2). This is in line with the fact that TDP-43 has been shown to be involved in multiple facets of RNA processing from RNA splicing, export and transport including those of synaptic transmission genes (Polymenidou et al., 2011). There was also an enrichment of genes related to Golgi vesicle transport and nucleocytoplasmic transport, which may be consistent with TDP-43's role in mRNA transport (Alami et al., 2014). Conversely, gene ontology analysis of significantly downregulated genes showed an enrichment of genes related to synaptic transmission, developmentally regulated genes and protein kinase activity. Some of these downregulated genes are involved in GABAergic synapses and are also dysregulated in mouse models of TDP-43 proteinopathies (Amlie-Wolf et al., 2015) (Table 3.3).

Weighted gene co-expression network analysis was used to determine whether we can identify modules of genes that correlate with each other. When this was performed on

significantly DEGs, a total of 22 modules significantly correlated with the loss of TDP-43 status. Using a module that was enriched for known processes related to neurodegeneration including ubiquitin processing, RNA splicing and DNA repair, a network was constructed using these genes (Fig 3.3). This network shows that genes related to these terms are interconnected and may correlate with each other. This was consistent with the role of TDP-43 in regulating RNA and ubiquitin dependent processes being important in misfolded protein clearance in neurodegeneration (King, Gitler, & Shorter, 2012), and raises the possibility that DNA damage may play a role in disease pathogenesis.

Loss of TDP-43 Is Associated with Alterations in Non-Coding Genic Elements

TDP-43 has been shown to preferentially bind long intronic RNA segments (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011). To supplement the above analysis based on whole gene expression, we sought to better understand whether loss of TDP-43 influences genic element (i.e. intron) expression. DEXSeq (Anders, Reyes, & Huber, 2012) was used to analyze the differential usage of genic elements including exons, introns and untranslated regions (UTR) irrespective of overall gene expression changes. There were a total of 5337 significantly differentially used genic elements linked to TDP-43 loss affecting 1337 genes (2487 upregulated elements and 2850 downregulated elements, Fig 3.2E). The dysregulated genic elements were annotated as either 5' UTR, exon, intron, or 3' UTR which showed an enrichment of both differentially used introns (76.37% vs 48%; chi-square test: $\chi 2=77.92$, p-val<0.0001) and 3' UTRs (13.40% vs 9.78%; chi-square test: $\chi 2=1713$, p-val<0.0001) in TDPneg nuclei compared to all annotated genic elements (Fig 3.2F). This was consistent with the fact that TDP-43 binds introns and 3' UTRs (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011; Tollervey et al., 2011). Given that many RNA binding proteins also bind to introns and 3' UTRs, we asked whether these elements were enriched for RNA binding proteins including hnRNPs. To address this, hnRNP and TDP-43 binding sites were determined using cross-linking immunoprecipitation sequencing data (Huelga et al., 2012; Tollervey et al., 2011) that was analyzed with PIPE-CLIP (B. Chen, Yun,

Kim, Mendell, & Xie, 2014). Indeed, there was a significant enrichment of genic elements bound by hnRNP, TDP-43 or both RBPs due to the loss of TDP-43 compared to all genic elements (16.56% vs 4.31%; chi-square $\chi 2=1837$; p-val<0.0001) (Fig 3.2G). These findings demonstrate that many of the changes that were predicted experimentally to occur due to TDP-43 dysfunction are indeed present in pathologic human neurons.

Transcriptome linked to TDP-43 loss highlights *TARDBP* autoregulation and selective vulnerability.

To further validate our sorting and sequencing methods, we sought to demonstrate that predicted changes associated with loss of TDP-43 activity could be observed, namely whether the loss of nuclear TDP-43 protein affects *TARDBP* autoregulation. TDP-43 is able to autoregulate by binding to a cryptic intron within its own 3' UTR that promotes intron splicing and usage of downstream polyadenylation sites resulting in a longer transcript isoform that is retained within the nucleus and subsequently degraded (Avendano-Vazquez et al., 2012; Ayala et al., 2011). The loss of nuclear TDP-43 protein is predicted to reduce *TARDBP* autoregulation, which should manifest as decreased splicing of the 3' UTR intron and a reduction of nuclear *TARDBP* transcripts with long 3' UTR tails. Indeed, visual inspection of coverage plots revealed several changes consistent with the reduction of *TARDBP* autoregulation including: (1) increased number of reads in 3' UTR intronic TDP-43 binding site in TDPneg nuclei, (2) increased number of reads in the extreme 3' UTR of *TARDBP* transcript and (3) reduced number of reads in TDPneg nuclei compared to TDPpos nuclei (Fig 3.4A). Quantification of the reads and junction reads spanning the 3' UTR intronic TDP-43 binding site show that TDPneg nuclei have significantly reduced junction reads (t-test, p=0.0014) and increased reads at the TDP-43 binding site (t-test, p=0.010) compared to TDPpos nuclei (Fig 3.4B-C), indicating decreased splicing of the cryptic 3' UTR intron in TDPneg nuclei as predicted. Moreover, there was a significant reduction in *TARDBP* reads, including reads within the extreme 3' UTR in TDPneg nuclei (t-test, p=0.0015) relative to

TDPpos nuclei (Fig 3.4D-E). Thus, loss of nuclear TDP-43 protein in pathologic human neurons leads to a loss of *TARDBP* autoregulation.

In addition to TDP-43 autoregulation, TDP-43 inclusion patterns within the frontal cortex have also been characterized (Lee et al., 2017) which has demonstrated that TDP-43 pathology is most abundant in superficial neocortical layers, as shown in Fig 3.4F.To determine whether this selective vulnerability was also reflected in the RNA-seq data, neuronal genes expressed in specific cortical layers were annotated and analyzed (Bernard et al., 2012; Darmanis et al., 2015; Hawrylycz et al., 2012; Molyneaux, Arlotta, Menezes, & Macklis, 2007; Zeisel et al., 2015). Out of all annotated neuronal genes, there was a significant enrichment of upper cortical layer (I-III) genes (chi-square test: $\chi2=26.36$, $p<0.0001$) and significant depletion of lower cortical layer (IV-VI) genes in TDPneg nuclei (chi-square test: $\chi2=25.69$, $p<0.0001$) (Fig 3.4G). Therefore, RNA sequencing of pathologic human neurons is capable of identifying the molecular fingerprints of selective vulnerability.

Loss of TDP-43 Is Associated with Global RNA Splicing Alterations.

As TDP-43 is known to regulate splicing of several transcripts, we tested whether splicing changes can be found within TDPpos and TDPneg nuclei. rMATS was used to identify alternative splicing changes using junction read counts (Shen et al., 2014). Loss of TDP-43 was associated with 1044 significant alternatively spliced events of which 49.62% were skipped exons, 22% were mutually exclusive exons, 15% were retained introns, 8.52% were alternative 3' splice sites, and 4.69% were alternative 5' splice sites (Fig 3.5A). When comparing TDPpos and TDPneg nuclei, there was no bias in terms of the number of inclusion vs. exclusion events for each of these alternative splicing categories (Fig 3.5B). Based on these and the above results, RNA sequencing of pathologic human neurons was able to demonstrate that the human disease indeed exhibits many of the predicted features of dysfunctional TDP-43 activity including massive gene expression and genic element expression changes, loss of *TARDBP* autoregulation, selective vulnerability of superficial neocortical neurons, and abundant splicing alterations.

Loss of TDP-43 is Associated with More Accessible Regions around LINE elements.

While we have validated many predicted features and functions of TDP-43, we wanted to utilize this method to discover novel disease mechanisms associated with TDP-43 loss. We have previously found that neuronal nuclei with TDP-43 aggregates in patients is associated with larger nuclei and that TDP-43 transgenic mice similarly show nucleomegaly and a profound loss of heterochromatin (Amlie-Wolf et al., 2015). Our RNA sequencing data above identified changes in genes involved in histone regulation, DNA repair and DNA damage (Table 3.2 and Fig 3.3). Thus, we hypothesized that TDP-43 loss may be associated with altered chromatin states. FACS was used to isolate neuronal nuclei for ATAC-seq to examine changes in chromatin accessibility in TDPpos and TDPneg neuronal nuclei from the same seven ALS/FTD patient samples used for RNA-seq. On average 31.1 million uniquely mapped single-end reads were obtained per sample after filtering and deduplication.

PCA was used to determine whether global changes in chromatin accessibility could be linked to the presence or absence of TDP-43. Principal component 1 and 3, which explain 25% and 9% of the variance within the ATAC-seq data, respectively, showed the highest correlation with TDP-43 status (Fig 3.6A). Principal component 2, which explain 14% of the variance, did not correlate with any tested experimental variable, including TDP-43 status, gender or date of extraction (data not shown). To quantify chromatin accessibility using the ATAC-seq data, non-overlapping genomic bins of 1,000 bp were created amongst the 22 human autosomes, resulting in a total of 2,875,012 bins. These genomic bins were used to obtain normalized ATAC-seq read counts genome-wide in order to ascertain whether there were genomic regions which were significantly more open/accessible due to loss of TDP-43. EdgeR was used to identify 3,457 significantly differentially accessible genomic regions (FDR < 0.05), of which 858 genomic regions were significantly more accessible and 2,599 regions were more closed in TDPneg nuclei (Fig 3.6B).

Differentially open/accessible genomic bins were annotated based on whether they were an annotated genomic element. Thus, regions were assigned based on the presence of gene-

associated elements (promoter, 5' UTR, intron, exon, 3' UTR) or an intergenic repeat elements (hereafter called "repeats"), with the remaining regions designated as non-repeat intergenic regions (hereafter called "intergenic") (Fig 3.6C-E). Interestingly, significantly differentially accessible genomic bins annotated as repeats were highly enriched in TDPneg nuclei in comparison to TDPpos nuclei (62.71 % vs. 29.45%, chi-square test: $\chi 2=368$, $p<0.0001$) and in comparison to all genomic regions (62.71% vs. 56.11%; $\chi 2=19.06$, $p<0.0001$). Differentially accessible regions associated with genes (promoters, UTRs, exon or intron) are significantly depleted in TDPneg nuclei compared to TDPpos nuclei (27.33% vs. 48.17%; chi-square test: $\chi 2=140$, $p<0.0001$). These regional annotations corroborate the above finding that repeat regions are significantly more accessible in TDPneg nuclei. Importantly, this suggests that loss of TDP-43 protein is associated with increased chromatin accessibility of intergenic repeat elements, which are typically heterochromatinized. This is consistent with an overall euchromatinization of pathologic neuronal nuclei.

To investigate the specific repeat family that may be associated with TDP-43 loss, repeat annotated bins were further subdivided into elements that have been annotated as repeat families using RepeatMasker annotations. These repetitive element families included long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), long terminal repeats (LTR), low complexity repeats (DNA), rolling circle (RC), satellite, or simple repeats amongst others. Differentially open/accessible regions associated with LINE repeats are significantly enriched in TDPneg nuclei compared to regions enriched in TDPpos nuclei (79.7% vs 37.1%; Fisher's exact test: $p<0.0001$) and compared to all genomic regions (79.7% vs 55.9%; chi-square test: $\chi 2=47.77$, $p<0.0001$) (Fig 3.6F-H). In contrast, differentially accessible regions associated with other repetitive elements are either not changed in TDPneg nuclei compared to TDPpos nuclei (SINEs: 5.2% vs 5.71%, Fisher's exact test: $p=0.80$) or were significantly decreased (LTRs: 8.5% vs 47.6%,Fisher's exact test: $p<0.0001$).

Having shown enrichment of LINE elements within genomic regions that are significantly more open/accessible in TDPneg nuclei, we expanded our analysis to determine whether TDP-43

loss was associated with increased chromatin accessibility of LINE elements genome-wide. Repeat associated genomic bins were annotated as LINE vs. non-LINE repeat elements, and ATAC-seq read distribution was compared between TDPneg and TDPpos nuclei (Fig 3.6I-J). TDPneg nuclei exhibit increased chromatin accessibility of 59.7% of LINE-associated genomic bins, which was significantly higher than other non-LINE repeat regions (48.9%; chi-square test: $\chi2=33,814$: $p < 0.0001$) and significantly higher than all genomic bins (Fig. 3.6K; 45.8%; chi-square test: $\chi2=106,486$: $p < 0.0001$). Therefore, these cell-type specific ATAC-seq analyses of post-mortem human brain suggest that the loss of TDP-43 protein is associated with enhanced chromatin accessibility around repetitive elements, specifically LINE elements.

LINE1 DNA content is increased in nuclei without TDP-43.

The LINE1 retrotransposon subfamily includes about 100 active LINE1 elements. Increased chromatin accessibility around LINE elements may permit reactivation of endogenous LINE1 elements, resulting in increased reverse transcriptase activity and possibly retrotransposition, which can be measured using qPCR for LINE1 DNA content. DNA from sorted neuronal nuclei were subject to qPCR quantification of LINE1 ORF2 DNA content, normalized to SATA satellite DNA content as a control repetitive DNA element which is not subject to retrotransposition or cell to cell variation. Paired analysis revealed that TDPneg nuclei exhibited significantly more LINE1 ORF2 DNA than TDPpos nuclei (Fig. 3.7A; paired t-test: $p=0.0262$). Thus, TDP-43 loss in post-mortem human brain is associated with increased LINE1 DNA content, raising the possibility that TDP-43 dysfunction may influence LINE1 retrotransposon.

Abnormal TDP-43 expression Increases Retrotransposition Activity.

Increased LINE chromatin accessibility and LINE1 DNA content in pathologic nuclei suggest that aberrant TDP-43 is linked to retrotransposition. To determine causality, TDP-43 was knocked out in cells and retrotransposition efficiency was determined. A retrotransposition assay was used where HeLa cells were transfected with a plasmid encoding LINE-GFP sequences

including (1) ORF1, an RNA binding protein required for retrotransposition, (2) ORF2, an endonuclease and reverse transcriptase required for retrotransposition, and (3) an anti-sense (reversed) sequence of eGFP harboring an internal intron. In cells transfected with this plasmid, GFP fluorescence is an indication of retrotransposition in that GFP protein can only be expressed if transcription, splicing, and retrotransposition occurs followed by transcription of GFP (in the reverse orientation) from the genomic integration site (Fig 3.7B). Cells that are analyzed by flow cytometry were gated to select for single cells and then for GFP expressing cells in normal HeLa cells (Fig 3.7C) and in L1 expressing cells (Fig 3.7D). GFP expression can only be detected in L1 expressing cells and when a retrotransposition event occurred. HeLa cells were transfected with this LINE-GFP plasmid together with Cas9 and 2 different guide RNAs specific to *TARDBP* to knockout TDP-43 expression. Indeed, CRISPR/Cas9 system was able to knockout TDP-43 in HeLa cells overexpressing the LINE-GFP plasmid evident by reduced protein expression via Western blot and TDP-43 loss via immunocytochemistry (Fig A1A-B). Interestingly, *TARDBP* knockout cells were associated with a significant increase in retrotransposition, as evidenced by increased GFP expressing cells compared to Cas9 expressing cells (Fig 3.7E; sgRNA 1 vs Cas9: intercept=12.33, β=2.67, p-val=0.0008; sgRNA 2 vs Cas9: β=2.20, p-val=0.0034). Importantly, control experiments where the LINE-GFP plasmid was replaced with a plasmid that drives GFP expression demonstrated that *TARDBP* knockout has no effect on GFP expression (Fig 3.7F).

While these results raise the possibility that the loss of nuclear TDP-43 increases retrotransposition, we considered the possibility that the stress of enhanced TDP-43 expression may also increase retrotransposition, given that TDPneg nuclei are derived from neurons harboring cytoplasmic TDP-43 aggregates. Indeed, when wildtype TDP-43 or TDP-43 lacking the nuclear localization signal (TDP-43ΔNLS) was coexpressed with the LINE expressing plasmid in HeLa cells, there was a significant increase in GFP positive cells compared to the pcDNA5/TO control vector, reflective of the finding that aberrant TDP-43 expression leads to increased retrotransposition (Fig 3.7G). A mixed effects linear regression model was used to demonstrate that increased retrotransposition was dependent on the overexpression of TDP-43 or TDP-

43ΔNLS (WT-TDP43 vs pcDNA: intercept= 7.04, β=2.57, p-val<0.0001; TDP-43ΔNLS vs pcDNA: β=1.24, p-val=0.0077). Importantly, when replacing the LINE-GFP plasmid with a GFP plasmid, TDP-43 overexpression had no effect on GFP expression (Fig 3.7H). Thus, abnormal TDP-43 expression contributes to increased retrotransposition, raising the possibility that TDP-43 pathology (both the loss of TDP-43 and aberrant TDP-43 expression) may enhance retrotransposons activity.

Discussion:

While experimental systems are invaluable for understanding disease mechanisms, experimental results need to be validated and understood in the context of human disease. In particular, cellular and animal models of TDP-43 proteinopathy do not fully recapitulate all the pathologic features of ALS or FTD. However, the cellular heterogeneity of the human brain hampers molecular analysis of ongoing neurodegenerative disease processes, requiring novel techniques to understand the molecular changes associated with human TDP-43 neuropathology. Here, we present a study using post-mortem human brain from ALS and FTD patients to understand the effects of TDP-43 pathology, including the functional consequences associated with the loss of nuclear TDP-43 in neurons. We developed a novel nuclei fractionation sorting method to isolate neuronal nuclei with and without TDP-43 to specifically identify the nuclear RNA and chromatin changes associated with TDP-43 pathology. Our analysis was able to confirm several hypotheses regarding the molecular consequences of TDP-43 nuclear clearance in the actual human disease including (1) massive gene dysregulation in pathways related to nucleocytoplasmic transport, synaptic transmission and RNA processing, (2) specific dysregulation of 3' UTR and intron segments, (3) abundant splicing changes, (4) altered *TARDBP* autoregulation, and (5) selective vulnerability of superficial neocortical neurons. Furthermore, we observed that TDP-43 pathology is associated with alterations in transcriptional networks related to DNA damage/repair, enhanced chromatin accessibility of LINE elements, and increased LINE1 DNA content. Finally, aberrant TDP-43 expression in cells increased retrotransposition activity. Collectively, these studies highlight the use of human brain fractionation techniques for deep molecular phenotyping, which is able to both confirm prevailing hypotheses regarding disease mechanisms and discover novel disease pathways that may contribute to neurodegeneration.

The cellular heterogeneity of the brain is a barrier in terms of understanding cell-specific molecular alterations. Because the human cerebral cortex has a ratio of 3.76:1 glia:neurons

(Azevedo et al., 2009), it is likely that much of the transcriptome changes observed from whole brain analyses come from glial cells. Indeed, a recent report showed that changes from whole brain RNA-seq from ALS, FTD, or AD patients represent changes in composition or abundance of microglial cell types and depletion of neuronal cell types, making it difficult to infer the molecular aberrations present within pathologic neurons (Srinivasan et al., 2016). Efforts to circumvent this have been employed including laser capture microdissection or single cell RNA-seq methods. However, these present technical difficulties in terms of either low yields or transcriptional noise (Ofengeim, Giagtzoglou, Huh, Zou, & Yuan, 2017; Srinivasan et al., 2016). Moreover, studying epigenetic or chromatin structure using laser-capture microdissection or single cell methods can be challenging. Importantly, gene expression changes observed from whole brain lysates represent an average of all cells, both diseased and normal, making it difficult to determine what changes are causal or consequential. Molecular changes from rare cells are often lost or underestimated using bulk sequencing approaches. This is particularly true in the context of this study, where TDP-43 negative nuclei only comprise 7% of all neuronal nuclei (<2% of all cells). As a result, it was imperative for us to use FACS on post-mortem brain to isolate neuronal populations.

While post-mortem brain is typically representative of end-stage disease, neurodegenerative diseases are incessantly progressive wherein post-mortem tissue can be conceptualized as a cross-sectional snapshot of disease. Thus, the neurodegenerative disease brain exhibits loss of cells which have already degenerated, but still encompasses (1) cell types that are relatively healthy, (2) cell types that are actively affected by disease-specific mechanisms, and (3) reactive, predominantly glial cell types that are responding to ongoing degeneration. Here, we have captured neuronal nuclei affected by disease (TDPneg) versus nuclei from relatively unaffected neurons (TDPpos) within the same tissue. As a result, we demonstrate that it is possible to bypass the complexities inherent to brain heterogeneity by isolating a specific subpopulation of pathologic neurons in order to identify the transcriptome differences linked to TDP-43 loss. Another important benefit is that since two neuronal

populations are derived from a single tissue specimen, paired statistical analyses can be performed which limits the patient-to-patient variability associated with human studies.

This method demonstrated that the major alterations that were presumed to result from the loss of TDP-43 based on experimental work are actually present in diseased human neurons. Given that TDP-43 is a RNA binding protein that functions broadly in RNA processing ranging from splicing, mRNA stability, and mRNA transport (Lee et al., 2012), the loss of TDP-43 was expected to lead to wide changes in the nuclear transcriptome. Within our data set, we were able to find thousands of dysregulated genes involved in RNA processing, transport, development and synaptic transmission pathways. These changes are similar to that observed using experimental models (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011). Some of these pathways may be activated to compensate for the loss of TDP-43, or may perhaps reflect the attempted reactivation of developmental or regenerative processes in pathologic neurons. Furthermore, we were able to validate that loss of TDP-43 preferentially affects 3' UTRs and intronic RNA segments (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011; Tollervey et al., 2011). Various splicing changes were identified in TDPneg nuclei, consistent with TDP-43's role in splicing. Altered *TARDBP* autoregulation was also observed in TDPneg nuclei as well, as predicted by various cellular or animal studies (Avendano-Vazquez et al., 2012; Ayala et al., 2011; Igaz et al., 2011). Importantly, demonstrating that all these predicted molecular changes can be seen in TDPneg nuclei serves as an important demonstration that our method of neuronal nuclei sorting is robust and able to capture salient features of TDP-43 pathology.

Having demonstrated that post-mortem fractionation of human brain can validate the predicted molecular aberrations associated with TDP-43 pathology, we sought to use our method to discover novel disease-associated pathways. In particular, chromatin remodeling and genomic structure has been implicated in aging and neurodegeneration. Oxidative stress has been linked to DNA damage that contributes to the chromatin changes observed during aging (Kirkwood, 2005). In the context of neurodegeneration, tau protein which forms neurofibrillary tangles in Alzheimer's disease has been shown to alter heterochromatin organization, increase DNA breaks

and consequently increase transcription of heterochromatic regions (Frost et al., 2014; Mansuroglu et al., 2016). Furthermore, previous data from our group has shown that neuronal nuclei with TDP-43 pathology from ALS/FTD patients exhibits nucleomegaly, and mice overexpressing cytoplasmic TDP-43 show nucleomegaly, loss of heterochromatin, and aberrant histone mRNA processing (Amlie-Wolf et al., 2015). All of these findings suggest that chromatin relaxation, particularly around repetitive elements, may be a fundamental phenomenon in aging or cellularly stressed cells. However, we wanted to investigate the potential role of TDP-43 in chromatin accessibility directly in both post-mortem tissue and cells. Our current RNA-seq study of TDPneg neuronal nuclei from human brain tissue revealed that significantly differentially expressed genes were enriched for transcripts that regulate histones. We also identified transcriptional subnetworks related to DNA damage and repair. This prompted us to perform ATAC-seq where we found that genomic regions harboring LINE elements were more accessible in TDPneg neuronal nuclei.

LINE elements are retrotransposable elements, which can be active in human genomes in both the germline and in somatic cells (Beck et al., 2010; Coufal et al., 2009; Erwin et al., 2014; Evrony et al., 2012). LINE and other repetitive elements are strictly heterochromatinized in order to suppress their expression and downstream genotoxic effects. While the damaging effects of retrotransposition is likely sufficient to induce neurotoxicity, LINE expression can be deleterious even in the absence of retrotransposition per se due to expression of ORF2 which has both reverse transcriptase and endonuclease activities. Increased LINE retrotransposition have been described in schizophrenia, Rett syndrome and ataxia telangiectasia (Bundo et al., 2014; Coufal et al., 2011; Muotri et al., 2010). Furthermore, activation of transposable elements has been observed in an age dependent fashion in *Drosophila* brain, contributing to memory decline and mortality (W. Li et al., 2013). Thus, LINE1 retrotransposition and transposable element activation appears to contributes to various human diseases.

With respect to ALS and FTD, TDP-43 has been shown to bind to repetitive elements and transposable elements (W. Li et al., 2012; Saldi et al., 2014) and has been postulated to be

involved in suppressing repetitive element activation (W. Li et al., 2012). Using a *Drosophila* model where human TDP-43 was overexpressed, LINE and LTR families were depressed (Krug et al., 2017), again supporting that TDP-43 is involved in suppressing repetitive elements. Interestingly, there is growing evidence supporting the reactivation of endogenous retroviruses and an increase in reverse transcriptase activity in ALS patients (W. Li et al., 2015; McCormick et al., 2008; Prudencio et al., 2017; Steele et al., 2005). Moreover, antiretroviral therapy has improved motor deficits in a subset of HIV patients who also have motor neuron and cognitive deficits (Bowen et al., 2016). The growing literature supporting a link between TDP-43 pathology and reactivation of endogenous genotoxic loci and our findings that loss of nuclear TDP-43 is associated with LINE accessibility, increased LINE1 DNA content and retrotransposition activity further substantiate the possibility that retrotransposon activity may contribute to disease.

In conclusion, we demonstrate here the successful implementation of a novel sorting method that circumvents the complexities associated with the molecular and cellular heterogeneity of the human brain. Many of the disease processes associated with ALS/FTD and TDP-43 proteinopathies were observed for the first time in the human brain, serving as an important validation of the various experimental discoveries regarding TDP-43 biology. Furthermore, beyond confirming pathways that were presumed to be occurring in the human brain, we uncovered new disease insights wherein we posit that TDP-43 pathology may be associated with increased transposon activity. The evidence for an increase in transposon activity stems from multiple additional sources including (1) the presence of altered DNA damage/repair transcriptional networks, (2) ATAC-seq data suggesting that TDP-43 pathology is associated with alterations in chromatin accessibility, (3) increased LINE-1 DNA content in TDPneg nuclei, and (4) enhanced LINE-1 retrotransposition in cultured cells with aberrant TDP-43 protein expression. Collectively, these results raise the possibility that pharmacologic inhibition of retrotransposon activity may mitigate the neurotoxic effects of TDP-43 pathology. Future studies that assess the extent of retrotransposition in ALS/FTD (such as single-cell sequencing to quantify whether retrotransposition is increased in TDPneg nuclei) and the relative

contribution of retrotransposon activity towards neurotoxicity may provide additional evidence as to whether pharmacologic inhibition of reverse transcriptase activity in ALS/FTD can have potential therapeutic benefit.

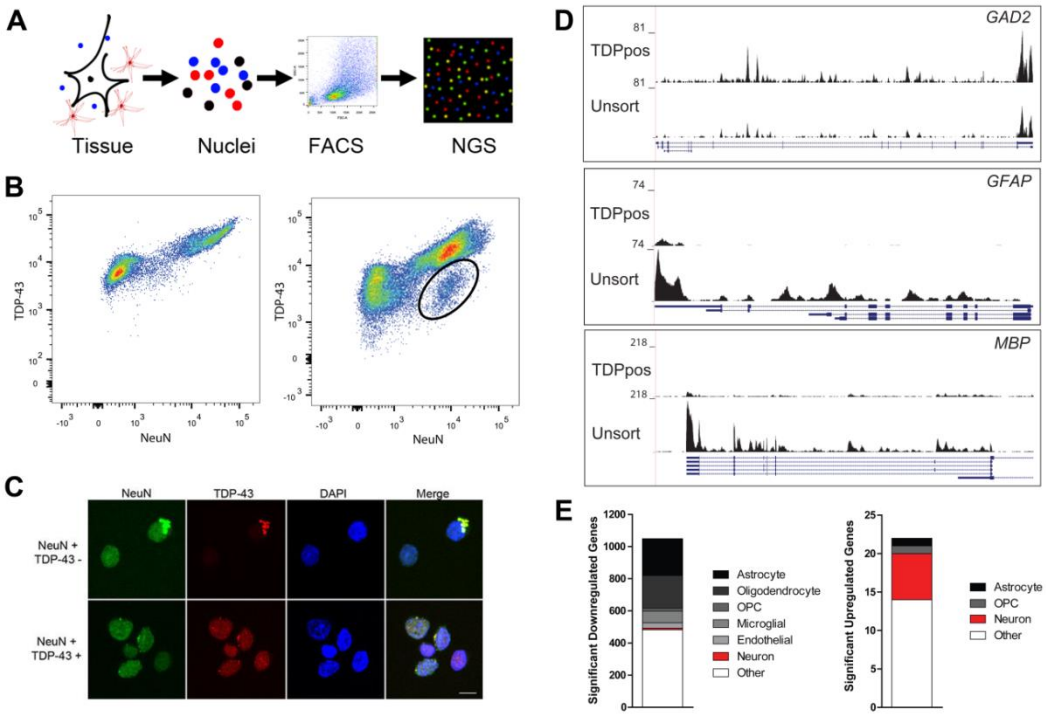

Figure 3.1: **Flow-Seq of post-mortem human brain was able to enrich for neuronal populations of interest**
(A) General schematic of Flow-Seq to isolate neuronal nuclei from post-mortem brain.
(B) Flow cytometry plots assessing TDP-43 fluorescence as a function of NeuN fluorescence of non-diseased post-mortem brain (left) and ALS/FTD patient brain (right). The non-diseased brain shows a non-neuronal population (NeuN negative, TDP-43 positive) and neuronal TDP-43 positive (NeuN positive, TDP-43 positive) population. The ALS/FTLD brain has an extra population (circled) of a neuronal population without TDP-43 (NeuN positive, TDP-43 negative).
(C) Confocal microscopy of sorted populations with and without TDP-43 assessing NeuN (green), TDP-43 (red) and DAPI (blue) fluorescence.
(D) Coverage plots of non-neuronal (*GFAP, MBP*) and neuronal (*GAD2*) genes in unsorted nuclei and neuronal nuclei with TDP-43.
(E) Enrichment of neuronal (red) genes within significantly upregulated genes (Fisher's exact test odds ratio=14.80; p-value=1.221E-5) and depletion of neuronal genes within significantly downregulated genes (Fisher's exact test odds ratio=0.37; p-value=0.00025) between sorted neuronal nuclei and unsorted nuclei.

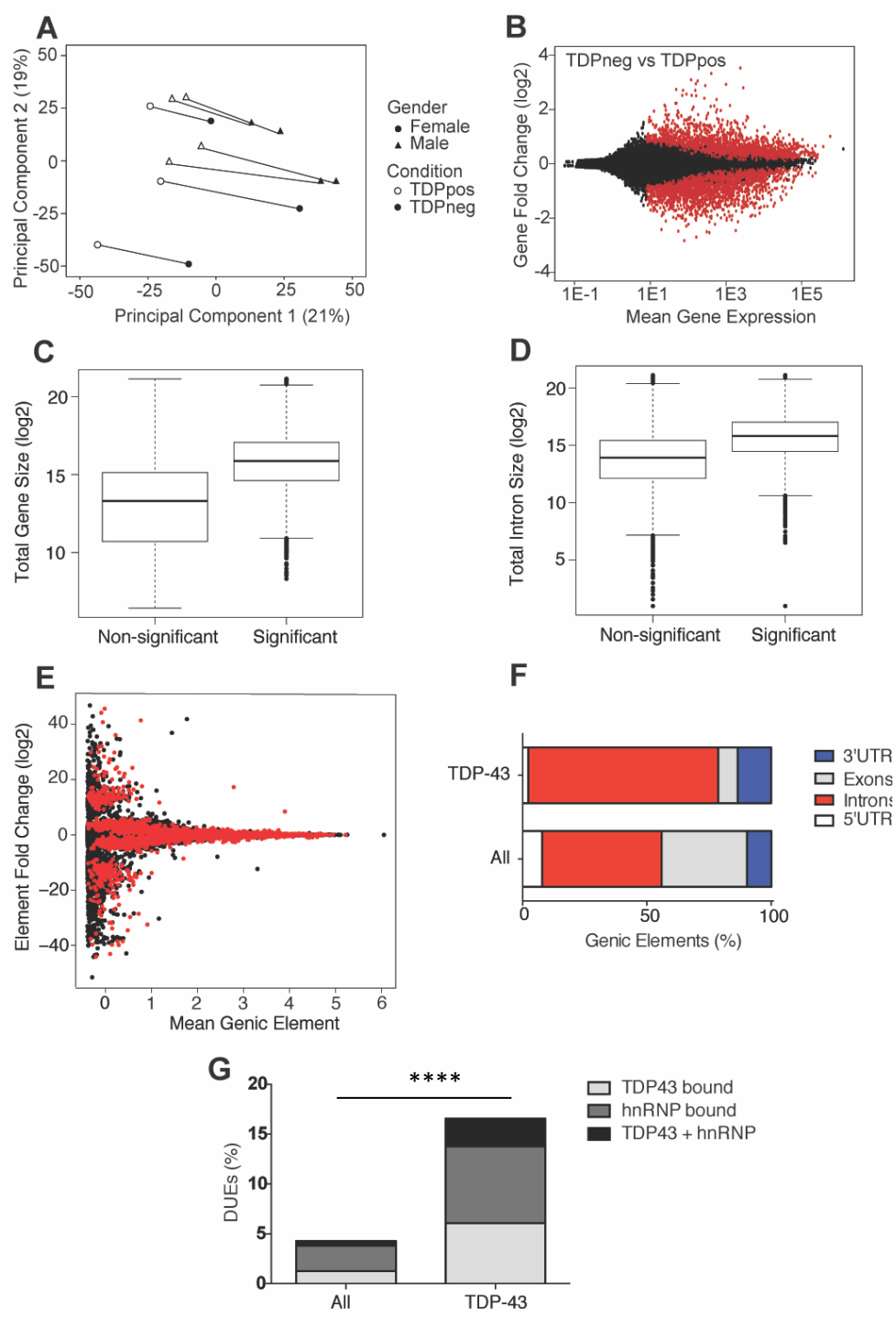

Fig 3.2: **Loss of TDP-43 is associated with massive gene expression changes and alterations in intron and 3' UTR usage**
(A) Principal component analysis of gene expression data with shape denoting gender and color denoting presence or absence of TDP-43. Lines are drawn to connect nuclei with and without TDP-43 from the same patient.

(B) MA plot of differentially expressed genes due to presence (TDPpos) or absence of TDP-43 (TDPneg) with red dots being significant differentially expressed genes (DEG) and black dots being expressed genes.
(C) Box plot of gene size (bp) in non-significant expressed genes and significant DEGs (t-test, $p<2.2E-16$).
(D) Box plot of gene size (bp) in non-significant expressed genes and significant DEGs (t-test, $p<2.2E-16$).
(E) MA plot of differentially used elements by DEXSeq upon TDP-43 loss. Black dots refer to non-significant differentially used elements and red dots are significantly differentially used elements upon loss of TDP-43.
(F) Differentially used genic elements from all expressed elements or those due to TDP-43 loss were as the 5' UTR, intron, exon or 3' UTR based on color. Chi-square test was performed to test enrichment of genic elements linked to TDPneg nuclei vs all genic elements (introns: 76.37% vs 48%, $\chi2=77.92$, p-val<0.0001; 3' UTRs: 13.40% vs 9.78%, $\chi2=1713$, p-val<0.0001).
(G) Genic elements were categorized based on presence of hnRNP or TDP-43 binding sites from publically available data. Chi-square test was performed to test enrichment of significantly differentially used elements that were bound by RBPs linked to TDPneg nuclei vs all genic elements with RBPs (16.56% vs 4.31%; $\chi2=1837$; ****p-val<0.0001).

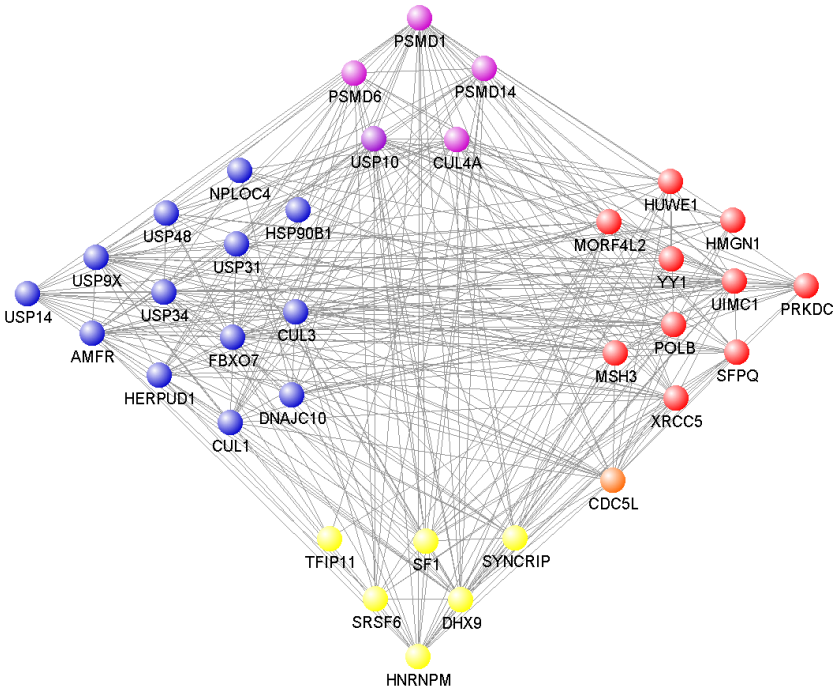

Fig 3.3: **Loss of TDP-43 is associated with genes that regulate RNA processing, DNA repair, DNA damage and proteostasis**
Network analyses of a WGCNA module derived from significantly differentially expressed genes linked to TDP-43 loss. Genes are categorized by gene ontology term in varying colors (blue = ubiquitin dependent catabolic processing, yellow = RNA splicing via transesterification, red = DNA repair; signal to DNA damage, purple = ubiquitin dependent catabolic processing and DNA repair/signal to DNA damage, orange = RNA splicing via transesterification and RNA splicing).

UPREGULATED GENES

		Term	Overlap	Adjusted P-value	Z-score
Transport	Biological Process	Golgi vesicle transport (GO:0048193)	55/202	0.000	-2.232
		intra-Golgi vesicle-mediated transport (GO:0006891)	17/33	0.000	-2.524
		establishment of protein localization to Golgi (GO:0072600)	10/16	0.004	-2.708
		protein targeting to Golgi (GO:0000042)	9/15	0.011	-2.618
		retrograde transport, vesicle recycling within Golgi (GO:0000301)	9/16	0.016	-2.632
		nuclear transport (GO:0051169)	46/209	0.019	-2.238
		nucleocytoplasmic transport (GO:0006913)	45/204	0.020	-2.224
		regulation of synaptic plasticity (GO:0048167)	35/120	0.001	-2.165
Synapse	Biological Process	synaptic transmission (GO:0007268)	84/434	0.011	-2.337
		regulation of excitatory postsynaptic membrane potential (GO:0060079)	18/52	0.014	-2.111
		regulation of postsynaptic membrane potential (GO:0060078)	19/58	0.017	-2.062
		regulation of synaptic transmission (GO:0050804)	50/245	0.032	-2.308
	Cellular Component	synaptic membrane (GO:0097060)	65/228	0.000	-2.334
		postsynaptic membrane (GO:0045211)	57/195	0.000	-2.292
		ionotropic glutamate receptor complex (GO:0008328)	18/49	0.001	-2.034
		synapse (GO:0045202)	46/243	0.048	-1.845
		axon (GO:0030424)	33/161	0.049	-1.735
		postsynaptic density (GO:0014069)	27/98	0.002	-2.122
		dendrite (GO:0030425)	53/236	0.001	-2.263
mRNA Processing	Biological Process	mRNA processing (GO:0006397)	78/397	0.012	-2.370
		regulation of RNA splicing (GO:0043484)	22/73	0.017	-2.113
		positive regulation of protein kinase activity (GO:0045860)	83/456	0.029	-2.401
		regulation of mRNA metabolic process (GO:1903311)	22/76	0.023	-2.144
		RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	39/177	0.031	-2.150
		regulation of mRNA processing (GO:0050684)	20/69	0.029	-2.103
		mRNA splicing, via spliceosome (GO:0000398)	39/177	0.031	-2.148
		RNA splicing, via transesterification reactions (GO:0000375)	40/184	0.032	-2.150
		mRNA transport (GO:0051028)	31/134	0.042	-2.181
		RNA transport (GO:0050658)	34/153	0.046	-2.169
		mRNA export from nucleus (GO:0006406)	19/69	0.050	-1.988
Organelles	Cellular Component	nucleolus (GO:0005730)	265/1653	0.000	-2.156
		Golgi membrane (GO:0000139)	59/308	0.012	-1.982

		Golgi apparatus (GO:0005794)	137/865	0.031	-1.936
Protein kinase	Molecular Function	protein serine/threonine kinase activity (GO:0004674)	93/449	0.000	-2.531
Histone Regulation	Molecular Function	lysine-acetylated histone binding (GO:0070577)	9/17	0.031	-3.162

Table 3.2: EnrichR GO terms of significantly upregulated genes due to TDP-43 loss.

DOWNREGULATED GENES

		Term	Overlap	Adjusted P-value	Z-score
Synapse	Biological Process	neuron projection guidance (GO:0097485)	66/367	0.010	-2.395
		axon guidance (GO:0007411)	66/367	0.010	-2.393
		synaptic transmission (GO:0007268)	74/434	0.019	-2.330
		gamma-aminobutyric acid transport (GO:0015812)	6/9	0.028	-2.208
	Cellular Component	transmembrane transporter complex (GO:1902495)	53/286	0.005	-2.220
		ion channel complex (GO:0034702)	49/258	0.005	-2.214
		transporter complex (GO:1990351)	53/291	0.007	-2.229
		synapse part (GO:0044456)	65/395	0.019	-2.241
		synapse (GO:0045202)	42/243	0.043	-2.002
		ionotropic glutamate receptor complex (GO:0008328)	13/49	0.046	-1.699
		clathrin-sculpted gamma-aminobutyric acid transport vesicle membrane (GO:0061202)	5/8	0.025	-0.190
	Molecular Function	ion channel activity (GO:0005216)	65/396	0.036	-2.447
		cation channel activity (GO:0005261)	49/285	0.036	-2.311
		gated channel activity (GO:0022836)	54/323	0.036	-2.303
		calcium channel activity (GO:0005262)	24/108	0.036	-2.294
		calcium ion binding (GO:0005509)	104/698	0.036	-2.284
Development	Biological Process	positive regulation of nervous system development (GO:0051962)	61/313	0.005	-2.414
		positive regulation of neurogenesis (GO:0050769)	57/283	0.005	-2.383
		positive regulation of cell development (GO:0010720)	63/358	0.020	-2.416
Protein Kinase	Molecular Function	transmembrane receptor protein kinase activity (GO:0019199)	23/82	0.011	-2.272
		transmembrane receptor protein tyrosine kinase activity (GO:0004714)	19/65	0.011	-2.164

Table 3.3: EnrichR GO terms of significantly downregulated genes due to TDP-43 loss.

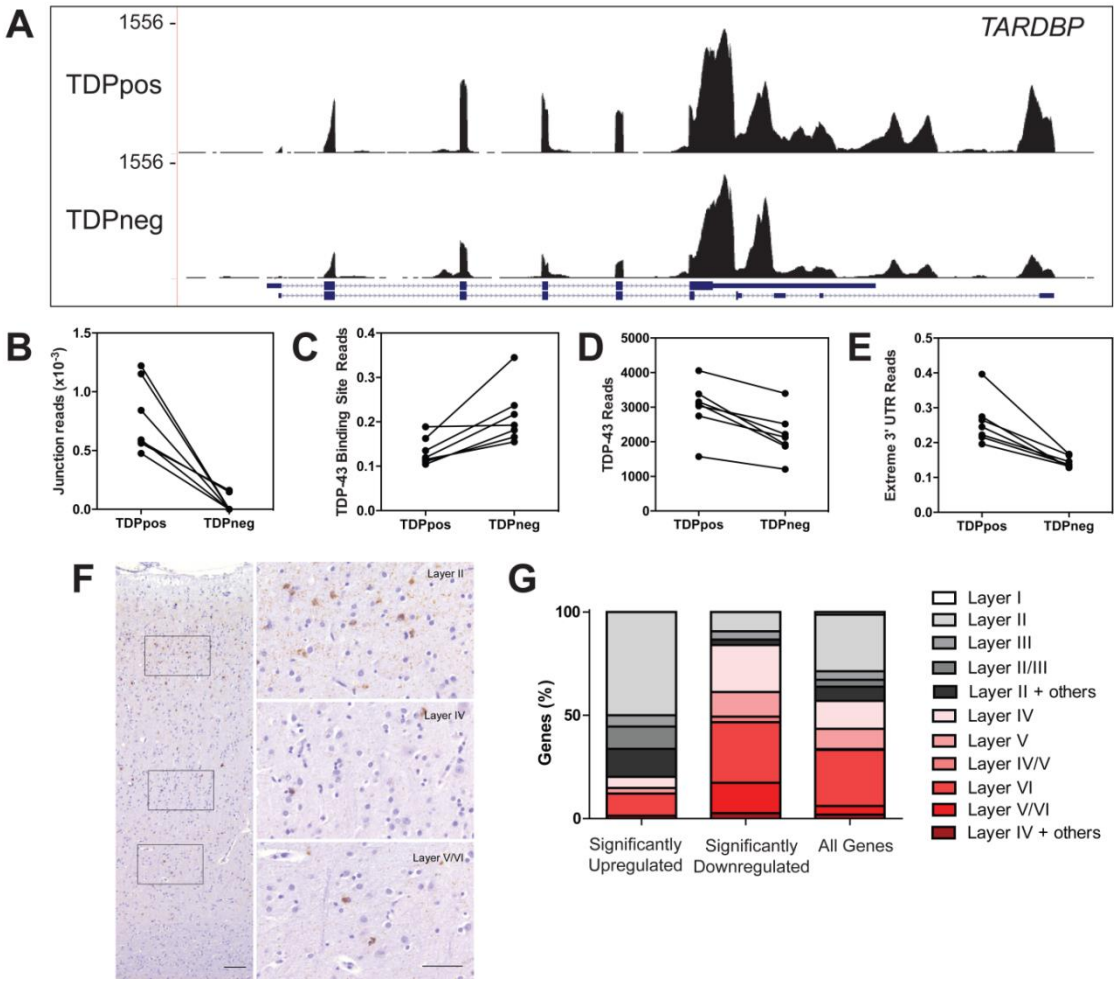

Fig 3.4: **Transcriptome linked to loss of TDP-43 highlights *TARDBP* autoregulation and selective vulnerability in human post-mortem brain**
(A) Coverage plot of *TARDBP* between TDPpos and TDPneg nuclei with the shorter and longer *TARDBP* isoform. There are more reads mapped to the longer 3' UTR in TDPpos nuclei compared to TDPneg nuclei.
(B) Junction reads mapped to the TDP-43 binding site were normalized to the total number of reads across *TARDBP* in TDPpos and TDPneg nuclei (t-test, p=0.0015).
(C) Reads mapping to the TDP-43 binding site were quantified and normalized to the *TARDBP* gene between TDPpos and TDPneg nuclei (t-test, p=0.01).
(D) Reads mapping to the gene *TARDBP* were quantified between TDPpos and TDPneg nuclei (t-test, p=0.0015).
(E) Reads within the extreme 3' UTR of the *TARDBP* gene were normalized to the reads across *TARDBP* in TDPpos and TDPneg nuclei (t-test, p=0.0018).
(F) Immunohistochemistry of TDP-43 aggregates found in an ALS/FTD patient with Type B pathology showing preferential TDP-43 pathology in superficial layers (Layer II) and little TDP-43 pathology in deeper cortical layers (Layer V/VI).
(G) A subset of significantly upregulated and downregulated genes were annotated according to cortical layers where red colors correspond to Layers IV-VI and black colors correspond to Layers I-III genes. Chi-square test was done to determine enrichment of upper cortical neurons within significantly upregulated genes in TDPneg nuclei ($\chi 2=26.36$, $p<0.0001$) and an enrichment of

lower cortical neurons within significantly downregulated genes ($\chi 2=25.69$, $p<0.0001$) in TDPneg nuclei.

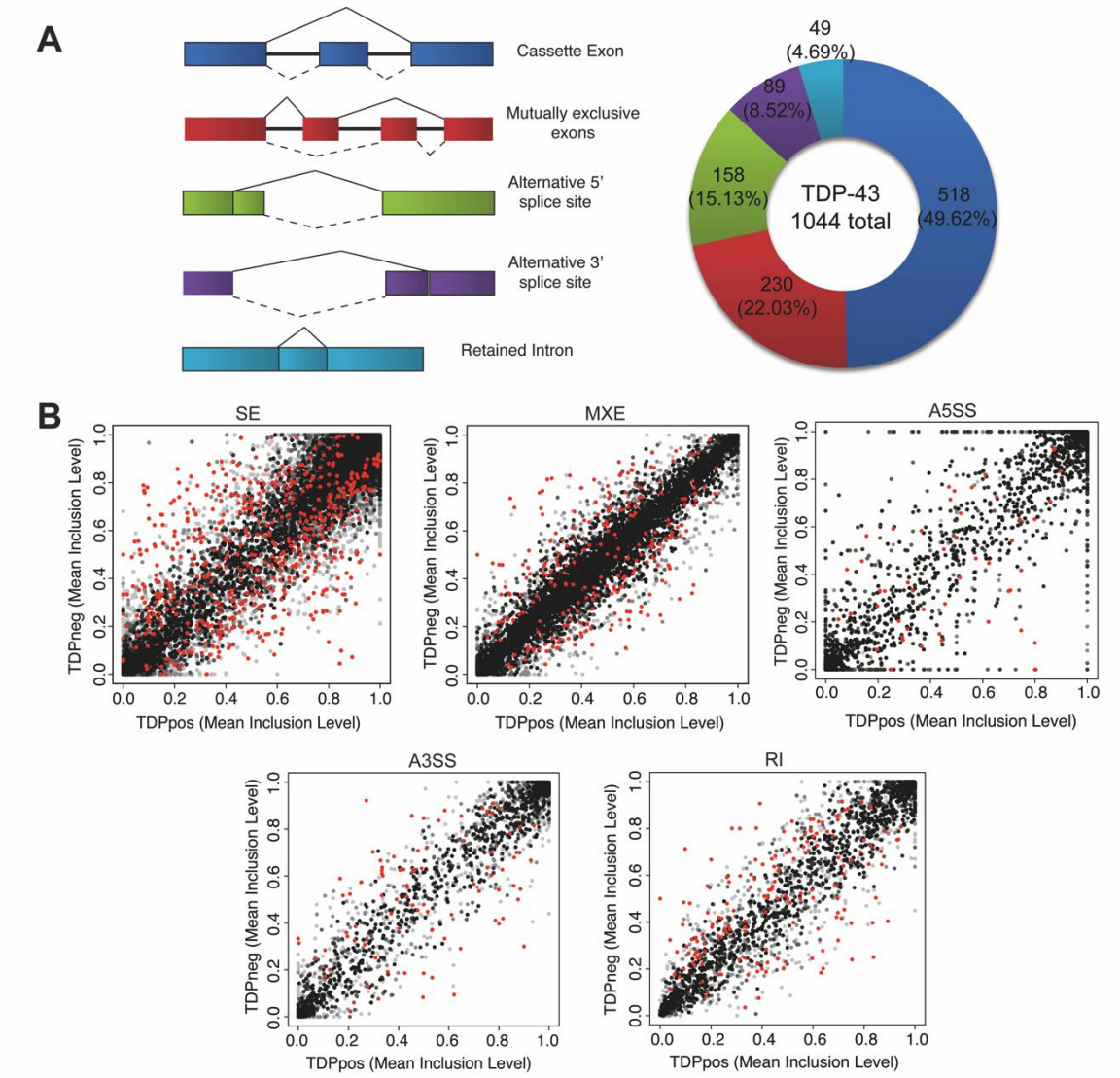

Fig 3.5: **Loss of TDP-43 is associated with massive alternative splicing changes**
(A) Representative examples of alternative splicing changes with color of each event denoted in the pie chart. Solid lines correspond to canonical splicing event and dotted lines correspond to alternative splicing event. The distribution of alternatively spliced events upon TDP-43 loss is quantified in the pie chart on the right.
(B) Mean inclusion levels of TDPneg nuclei was plotted against mean inclusion levels of TDPpos nuclei for each queried alterative splicing event with significant alternative splicing events in red. The darker a dot is, the more junction read counts there were for a particular splicing event.

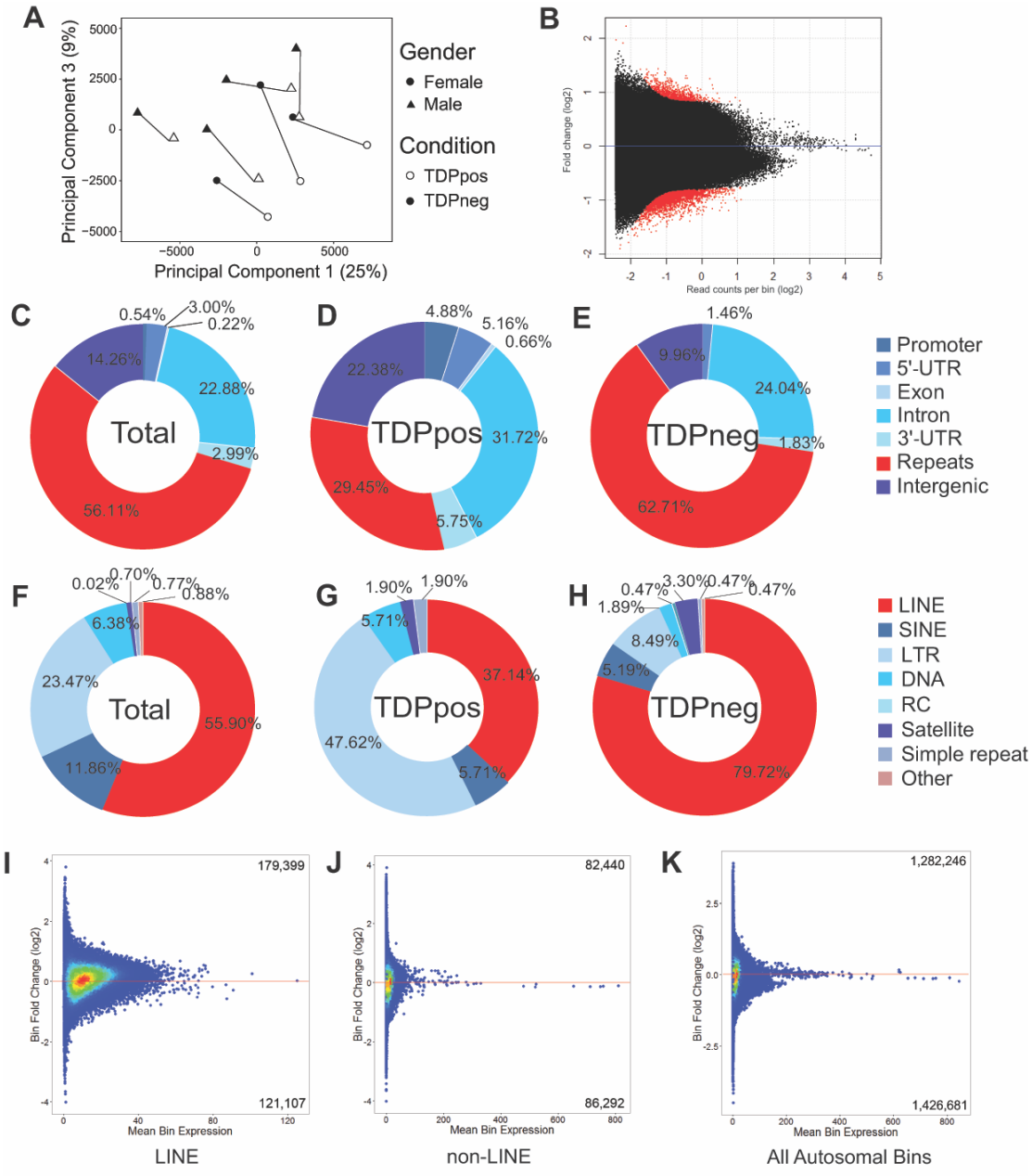

Fig 3.6: **LINE elements are more accessible in post-mortem human neuronal nuclei without TDP-43.**
(A) Principal component analysis of ATAC-seq bin expression with shape denoting gender and color denoting condition with a line drawn for each patient.
(B) MA plot of bin expression changes linked to TDP-43 loss with red dots being significantly differentially expressed bins and black being non-significant bins.
(C-E) Percentage of bins mapped to genic elements (promoter, 5‘ UTR, exon, intron, 3‘ UTR) and mapped to repeat regions or intergenic regions in all autosomal bins (C), significantly open bins in TDPpos nuclei (D) or significantly open bins in TDPneg nuclei (E). Fisher’s exact test was

performed to determine enrichment of elements in TDPneg vs TDPpos nuclei (repeats: 62.71% vs 29.45%, Fisher's exact test: $\chi 2=368$, $p<0.0001$) or compared to all autosomal bins (repeats: 62.71% vs 56.11%, Fisher's exact test: $\chi 2=19.06$, $p<0.0001$).
(F-H) Percentage of bins annotated as repeat elements that belong to repeat families such as long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), long terminal repeats (LTR), low complexity DNA (DNA), rolling circle (RC), satellite repeats, simple repeats or other (rRNA, snRNA, srpRNA) in UCSC RepeatMasker tracks in all autosomal bins (F), significantly open bins in TDPpos nuclei (G), or significantly open bins in TDPneg (H). Fisher's exact test was done to determine enrichment of LINE elements in TDPneg vs TDPpos nuclei (LINE: 79.7% vs 37.1%, Fisher's exact test: $p<0.0001$) and in TDPneg vs all bins (79.7% vs 55.9%, Fisher's exact test: $p<0.0001$).
(I-K) MA plot using bin fold change linked to TDP-43 loss and mean bin expression of LINE elements (I), non-LINE repeat elements (J), and all autosomal bins (K).

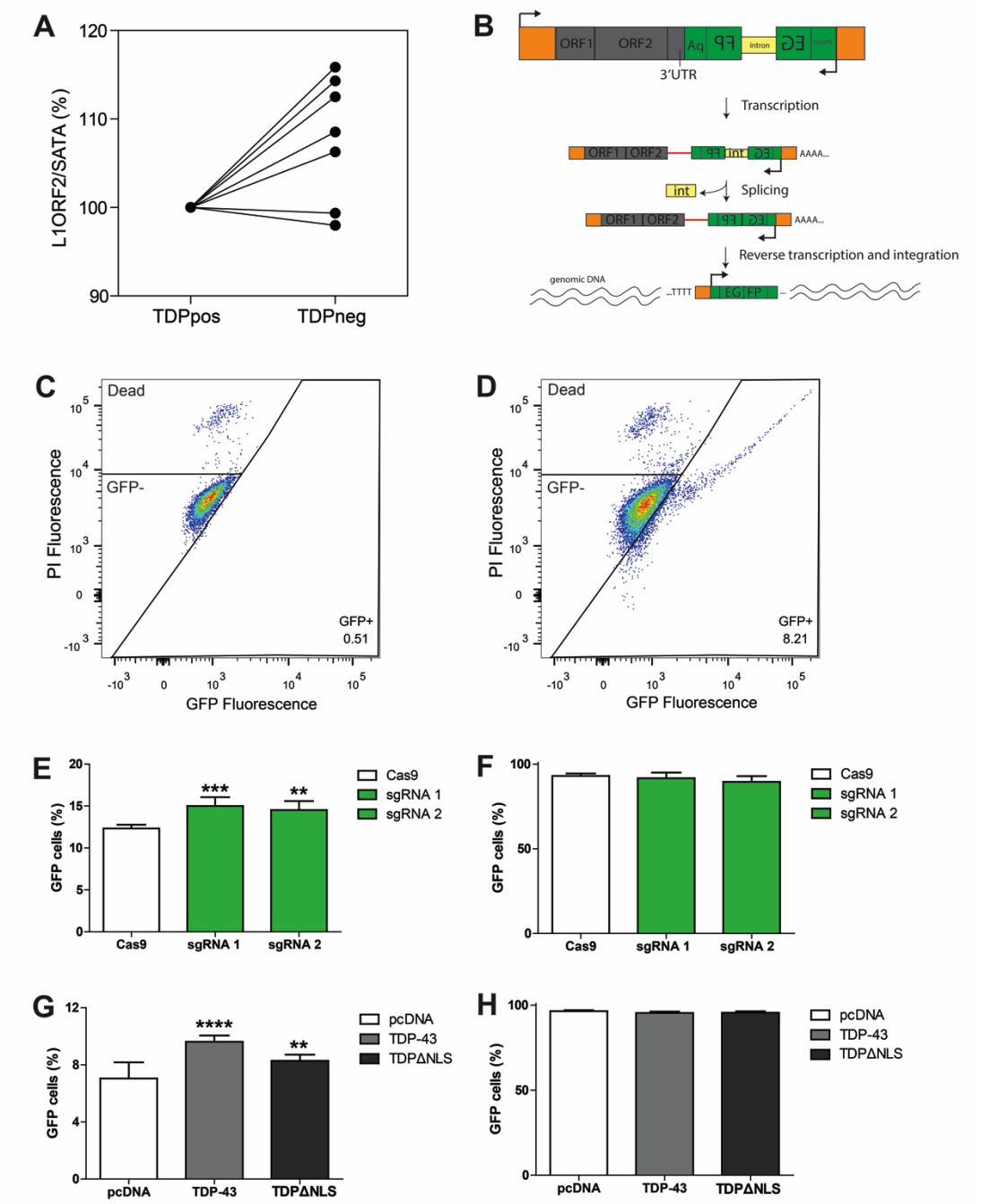

Fig 3.7: **Loss of TDP-43 is associated with increased retrotransposition**
(A) L1ORF2 DNA normalized to SATA repeat DNA in TDPpos and TDPneg neuronal nuclei from post-mortem human brain (n=7) (t-test, p=0.0262).
(B) Retrotransposition assay using L1 expressing plasmid containing (1) ORF1 (RNA binding protein), (2) ORF2 (endonuclease and reverse transcriptase), (3) reverse sequence of EGFP harboring an internal intron under the CMV promoter. After transcription of this plasmid, the intron flanking the EGFP sequence would be spliced out. After reverse transcription and integration of the DNA into genomic DNA, only cells that have undergone retrotransposition would be expressing EGFP and can be detected by flow cytometry.

(C-D) Representative flow cytometry plots of untreated cells(C) and L1 expressing cells (D) showing PI fluorescence as a function of GFP fluorescence.
(E) Retrotransposition efficiency as measured by percentage of GFP expressing cells in cells overexpressing L1 plasmid and either (1) Cas9 plasmid, (2) 2 different guide RNA (sgRNA 1 and sgRNA 2) that target *TARDBP* for CRISPR/Cas9 gene editing (n=2). A mixed effects linear regression model was used to show that knocking out TDP-43 leads to increased retrotransposition compared to pcDNA overexpression (sgRNA 1 vs Cas9: intercept=12.33, β=2.67, p-val=0.0008; sgRNA 2 vs Cas9: β=2.20, p-val=0.0034).
(F) Retrotransposition efficiency as measured by percentage of GFP expressing cells in cells overexpressing GFP expressing plasmid and Cas9 or *TARDBP* targeting guide RNAs (n=2). There was no significant difference in GFP expressing cells.
(G) Retrotransposition efficiency as measured by percentage of GFP expressing cells in cells overexpressing L1 plasmid and pcDNA5/TO, wildtype TDP-43 (TDP-43) or TDP-43 with 2 mutated nuclear localization signals (TDP-43ΔNLS) (n=4). A mixed effects linear regression model was used to show that overexpressing WT-TDP43 and TDP-43ΔNLS leads to increased retrotransposition compared to pcDNA overexpression (WT-TDP43 vs pcDNA: β=2.57, p-val<0.0001; TDP-43ΔNLS vs pcDNA: β=1.24, p-val=0.0077).
(H) Retrotransposition efficiency as measured by percentage of GFP expressing cells in cells overexpressing GFP expressing plasmid and pcDNA5/TO, TDP-43 and TDP-43ΔNLS (n=4). There was no difference in GFP expressing cells.

Contributions and Acknowledgements:

Elaine Y Liu designed the experiments, developed the Flow-Seq method, performed the RNA-seq experiment, analyzed the RNA-seq data, performed and analyzed the retrotransposition experiments and prepared the paper.

Jenny Russ designed the experiments, developed the bioinformatics pipeline analyses, performed the ATAC-seq, analyzed part of the ATAC-seq results, performed the LINE1 qPCR assay, and prepared the paper.

Alexandre Amlie-Wolf analyzed part of the ATAC-seq results and contributed to the paper.

Edward B Lee designed the experiments, analyzed results and edited the paper.

I would like to thank the University of Pennsylvania Flow Cytometry Core, particularly Paul Hallberg for doing the FAC sorting and the University of Pennsylvania Next Generation Sequencing Core.

CHAPTER 4:

C9orf72 Transcriptome is Associated with Loss of C9orf72 Protein.

Abstract:

A hexanucleotide G_4C_2 repeat expansion in *C9orf72* is the most common genetic cause of familial and sporadic cases of amyotrophic lateral sclerosis and frontotemporal degeneration. The mutation is linked to loss of C9orf72 protein and accumulation of toxic RNA and dipeptide repeat aggregates. The accumulation of toxic RNA is thought to sequester RNA binding proteins, which in turn, can alter RNA processing. Previous transcriptome studies have shown that the *C9orf72* repeat expansion is linked to large splicing alterations and transcriptome changes, which supports that the repeat expansion contributes to disease via altered RNA processing. Here, I used a subcellular fractionation method and FACS to enrich for normal neuronal nuclei from post-mortem human ALS/FTD brains. In this paradigm, neuronal nuclei with TDP-43 pathology, which is observed in all *C9orf72* expansion carriers, was depleted. I show that the *C9orf72* expansion is associated with mild gene expression changes. Dysregulated genes were enriched for vesicle and lysosomal involvement, which is consistent with C9orf72 protein function. The *C9orf72* transcriptome can also be explained by the depletion of pathologic TDP-43 nuclei. Further transcriptome analysis shows that the molecular signature linked to the *C9orf72* expansion is associated with a loss of C9orf72 protein rather than gain of toxic RNA or dipeptide repeat pathology. Thus, I show that the *C9orf72* transcriptome is shaped by the depletion of pathologic nuclei and loss of *C9orf72* protein. Furthermore, this work suggests that while RNA binding protein sequestration may not be the main factor to *C9orf72* mediated toxicity.

Introduction:

Amyotrophic lateral sclerosis (ALS) and frontotemporal degeneration (FTD) are two disparate neurodegenerative diseases that have many genetic similarities. ALS is a motor neuron disease that primarily affects the upper and lower motor neurons in the motor cortex and spinal cord, respectively, whereas FTD affects the frontal and temporal lobes, thereby affecting cognition. Recently, the most common genetic cause of familial and sporadic cases of ALS and FTD was identified as a hexanucleotide G_4C_2 repeat expansion found in the gene *C9orf72* (DeJesus-Hernandez et al., 2011; Renton et al., 2011). The repeat expansion is found in the first intron of *C9orf72* and can be transcribed into pre-mRNA that contains the repeat. C9orf72 protein has been predicted to be a DENN Rab GTPase, and shown to be involved in endosomal and lysosomal trafficking (Aoki et al., 2017; D. Zhang, Iyer, He, & Aravind, 2012). Three potential mechanisms of toxicity have been implicated to contribute to disease (1) haploinsufficiency and subsequent loss of protein, (2) RNA foci and titration of RNA binding proteins, and (3) dipeptide repeat protein aggregates (reviewed in S. C. Ling et al., 2013). It is currently unclear which mechanism contributes most to toxicity.

Gain of toxic function via RNA foci or dipeptide repeat proteins have been seen in cellular and animal models. Indeed, many RNA binding proteins have been shown to colocalize with repeat containing RNA as RNA foci, suggesting that RNA binding protein titration may be mediating toxicity (Cooper-Knock et al., 2014; Y. B. Lee et al., 2013). Overexpression models of the dipeptide repeat proteins in both cell and animal models have shown that the repeat expansion and dipeptide proteins confer toxicity via nucleolar stress and nucleocytoplasmic transport defects (Freibaum et al., 2015; Jovicic et al., 2015; May et al., 2014; Tao et al., 2015; Wen et al., 2014). Both examples show that gain of toxic function may be altering cellular homeostasis. Furthermore, halting the expression of the G4C2 repeat using antisense oligonucleotides, thereby preventing mutant RNA from being transcribed, reduced RNA foci pathology (Donnelly et al., 2013; Lagier-Tourenne et al., 2013). These all point to the idea that *C9orf72* pathology may be toxic. However, RNA foci and dipeptide repeat proteins do not

correlate with regions of degeneration (Ash et al., 2013; DeJesus-Hernandez et al., 2017; Mackenzie et al., 2015), raising questions about the relevance of whether this pathology contributes to disease.

On the other hand, loss of C9orf72 protein has been substantiated by the fact that *C9orf72* expansion carriers have reduced *C9orf72* mRNA and protein expression (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Waite et al., 2014b). Thus, it is possible that loss of C9orf72 protein may be toxic. Indeed, an animal model where *C9orf72* was depleted in zebrafish causes motor deficits (Ciura et al., 2013b). However, knockout mouse models of *C9orf72* do not show neuronal defects but rather, exhibit an inflammatory phenotype (Burberry et al., 2016; Koppers et al., 2015; O'Rourke et al., 2016). Furthermore, we have shown that a subset of *C9orf72* expansion carriers are hypermethylated, and this is associated with reduced C9orf72 associated pathology, later age at death in FTD patients, and improved cognitive skills (E. Y. Liu et al., 2014; McMillan et al., 2015; Russ et al., 2015). Consequently, it is possible that loss of C9orf72 protein may not be as detrimental as presumed and further makes it unclear which mechanism of toxicity contributes most to disease.

Recent efforts to try to reconcile this dilemma have shifted efforts to transcriptomic studies to determine whether RNA dysregulation can contribute to neurodegeneration. Indeed, frontal cortical and motor neuron transcriptome analysis from sALS patients bearing the repeat expansion showed massive splicing changes, dysregulation of RNA processing pathways, and widespread alternative polyadenylation in *C9orf72* expansion carriers (R. H. Batra, K.; Vu, A.;Rabin, S.J.; Baughn, M.W.; Libby, R.T.; Hoon, S.; Ravits, J.; Yeo, G.W., 2016; Cooper-Knock et al., 2015; Prudencio et al., 2015). However, these studies either simplify the composition of the brain by identifying gene expression changes that are likely coming from glial cells rather than neurons. On the other hand, laser capture microdissected tissue generally has low yields and has many preprocessing steps that can affect RNA quality (Curran, McKay, McLeod, & Murray, 2000).

Thus, alternative efforts to understand the molecular changes linked to the *C9orf72* expansion should be pursued.

We have previously reported the use of fluorescence activated cell sorting (FACS) to identify transcriptome signatures in neuronal nuclei without TDP-43 from post-mortem brain in ALS/FTD patients. Here, we have used the same method to determine the role of the *C9orf72* repeat expansion in post-mortem neurons from ALS/FTD patients. We found mild gene expression changes with dysregulated genes being involved in vesicle and lysosomal involvement and alternative splicing. The repeat expansion is associated with mild splicing changes, and our transcriptome analyses reveals that the molecular signature is associated with a loss of C9orf72 protein rather than gain of toxic RNA and dipeptide repeat pathology.

Materials and Methods:

Clinical and pathologic assessment.

Human autopsy tissue was obtained from the University of Pennsylvania Center for Neurodegenerative Disease Research Neurodegenerative Disease Brain Bank as described (Toledo et al., 2014).

Human brain nuclei isolation and FACS.

Isolation was performed exactly as in Chapter 3. Briefly, mid-frontal neocortex of all cases were dounce homogenized using pestil B (Kimble Chase) in 0.25M sucrose and adjusted to final molarity of 1.6M sucrose in TKM. The homogenate was spun on a 1.8M sucrose cushion on the Beckman Coulter XPN-80 ultracentrifuge at 40,000g for 40 minutes at 4^{o}C (Beckman Coulter Inc, Indianapolis, IN, USA). Isolated nuclei were stained with Alexa Fluor 647 conjugated to 2089 **(**rabbit polyclonal C-terminal anti-TDP-43 antibody, Center for Neurodegenerative Disease Research, University of Pennsylvania), Alexa Fluor 488 conjugated NeuN (EMD Millipore, Billerica, MA, USA), and DAPI (Invitrogen, Carlsbad, CA, USA)**.** Alexa Fluor 647 was conjugated to 2089 according to the APEX Alexa Fluor 647 Antibody labeling kit protocol (Thermo Fisher Scientific, Waltham, MA, USA). Stained nuclei were sorted for single cells containing TDP-43 and NeuN on the BD FACSAria II (BD Biosciences, San Jose, CA, USA) at 20 psi on 100µm nozzle.

RNA Isolation and RNA-seq Library Generation.

Isolation and library generation was done as in Chapter 3. RNA was extracted using the standard protocol within the AllPrep DNA/RNA Micro kit (Qiagen, Germantown, MD, USA). RNA quality from sorted nuclear RNA was determined based on Bioanalyzer Picochip analysis (Agilent, Santa Clara, CA, USA). Isolated RNA was amplified, made into cDNA, sheared, and

libraries were made. The library was quantified using the Qubit dsDNA kit (Invitrogen) and Kapa library quantification kit (KapaBiosystems, Boston, MA). cDNA libraries were pooled, clustered on the cBot and subject to 100 or 125 base pairs paired end reads on the HiSeq 2000 or 2500 (Illumina, San Diego, CA, USA).

Pre-processing, mapping and filtering of RNA-seq data.

RNA-seq analysis was done as in Chapter 3. Briefly, raw sequencing reads were demultiplexed through the UPenn Functional Genomics Core and analyzed for quality control using FastQC. Reads were mapped to the human genome (GRCh38, GENCODE release 22) using STAR and only uniquely mapping reads were selected for further analysis. Ribosomal and mitochondrial reads were removed and SAM files were converted to BAM files using samtools view and BAM files were sorted by coordinate with samtools sort.

Creation of non-overlapping gene, exon and intron annotations.

Annotations were generated as in Chapter 3. Briefly, annotations were based on the comprehensive gene annotation file of the GENCODE Release 22 (GRCh38.p2). GTF file was loaded into R (Version 3.2.2; R Core Team (2015): “R: A Language and Environment for Statistical Computing”, R Foundation for Statistical Computing, Vienna, Austria) and converted into a TranscriptDb object. From the TranscriptDb object, all annotated Ensembl genes and their exons were pulled out using exonsBy (by=”gene”) and Ensembl gene IDs were replaced by official gene symbols with biomaRt (Version 2.26.1). Genes and introns were defined as previously described in the above chapter. Regions shared by overlapping genes were removed to count reads that map to one gene or to genic elements (exon or intron) from one gene. Exons and introns annotations were used for subsequent analyses in the R package DEXSeq (1.16.10).

Differential gene and genic element expression analysis.

The genomic annotation used for read counting was done exactly as in the genomic annotation from Chapter 3. The mapped, filtered RNA sequencing reads were counted using a custom R script including the R packages Rsamtools (1.22.0), GenomicFeatures (1.22.8) and GenomicAlignments (1.6.3). Briefly, sorted BAM files were loaded into R (3.2.2) and the number of reads mapping to genes, exons or introns, was computed.

Genes were analyzed for differential expression using the R package DESeq2 (1.10.1). Analysis was used to determine differences between nuclei from both *C9orf72* expansion carriers and non-diseased controls (option design=~experiment in DESeqDataSetFromMatrix tool where experiment is the subject id). Gender effects were removed with the R package sva (3.18.0) according to the DESeq2 vignette. Manual annotation of gene descriptions were performed using gene descriptions on NCBI.

Differential genic element expression analysis was done using DEXSeq (1.16.10). Briefly, the sorted BAM files were used to count the number of reads mapping to exons and introns using the exon and intron annotation generated above and computed using findOverlaps and countSubjectHits. Reads mapping to exon-intron junctions were excluded. Gender was controlled for by including it in the linear model used in the analysis by adding the term "gender:exon". The following linear models were used full model = ~sample + exon + gender:exon + experiment:exon and reduced model = ~sample + exon + gender:exon + experiment:exon and reduced model = ~sample + exon + gender:exon, where sample is the sample id and experiment the subject id. In addition, dispersions were estimated using the tool estimateDispersions with option fitType='local'. Changes in expression were significant if Bonferroni-Hochberg multiple testing adjusted p-values were less than 0.05.

Alternative splicing analysis.

Splicing analysis was performed as in Chapter 3. Briefly, all FASTQ files were trimmed to 100bp, aligned to GRCh38 using STAR; ribosomal and mitochondrial mapped reads were removed. SAM files were converted to sorted BAM files and rMATS.3.0.9 was run using default

parameters with the following options (-t paired –len 100 –c 0.05 – analysis U). Significant alternative splicing events were used if Bonferroni-Hochberg multiple testing adjusted p-values were less than 0.05.

RNA binding protein CLIP analysis.

CLIP analysis was done as in Chapter 3 using published TDP-43 iCLIP data from SH-SY5Y cells (Tollervey et al., 2011) and hnRNP A, A2B1, F, M, U CLIP data from HEK293 cells (Huelga et al., 2012). Briefly, hg18 Bowtie files were converted to FASTQ files, and aligned to hg38. PIPECLIP was run using the python script. Using R, the "GenomicRanges" package with 'findOverlaps' option was used to determine which bins had RNA binding protein sites. Chi-square analysis was done to determine whether there was a significant enrichment of bins with RNA binding protein sites within the significantly differentially used bins linked to the *C9orf72* mutation.

Principal Component Analysis.

Principal component analysis was done using the log transformed read count dataset in R with the 'stats' package and 'prcomp' function. The coordinates were retrieved and used to plot principal component 2 (PC2) vs principal component 1. To determine the relationship between PC2 vs *C9orf72* expression, the coordinates of PC2 for each sample was plotted against the normalized *C9orf72* expression calculated from DESeq2 from each sample. Pearson's correlation was performed to determine the correlation between PC2 and *C9orf72* expression. Gene ontology analysis of the top 1% of all genes that contribute to PC2 was done using Webgestaldt (J. Wang, Duncan, Shi, & Zhang, 2013) with the overrepresentation enrichment analysis using the 'geneontology' functional database with a FDR corrected significance level < 0.05.

Methylation Correlation.

Methylation levels for the *C9orf72* promoter in patients was determined as previously described (Russ et al., 2015). Briefly, genomic DNA from the cerebellum was extracted using the Qiagen DNeasy Blood and Tissue kit and subject to overnight digestion with HhaI and HaeIII (double-digested) or just HaeIII (mock) alone. A small aliquot of DNA was amplified using primers flanking the HhaI cut site within the *C9orf72* promoter region using 2x FastStart SYBR Green Master (Roche) on the ABI StepOnePlus machine. The difference in cycles to threshold amplification between double and mock digested DNA was calculated as methylation values. Spearman's correlation was calculated using gene counts for each gene and methylation values for each *C9orf72* mutation carrier. R package 'lsr' with 'correlate' function (with options corr.method="spearman" and p.adjust.method="fdr") using the Spearman's correlation and FDR adjusted p-values was used. Only correlations from genes with HUGO gene symbols were calculated and plotted against gene fold change for each gene that was calculated by DESeq2. This was done using both significantly differentially expressed genes linked to the C9orf72 mutation (DESeq2 FDR p-value < 0.05) and genome-wide using all expressed genes.

Case	Sex	Age of Death	Age of Onset	PMI	Diagnosis	MND	Dementia	RIN frontal
C9orf72	F	71	62	3.5	FTLD	N	Y	6.4
C9orf72	M	57	55	8.5	FTLD-ALS	Y	Y	7.7
C9orf72	F	61	57	12	FTLD	N	Y	6.9
C9orf72	F	73	N/A	N/A	FTLD	N	Y	8.2
C9orf72	M	77	71	5	FTLD	N	Y	6.2
C9orf72	M	57	55	6	FTLD-ALS	Y	Y	7.3
C9orf72	M	75	71	10	FTLD-ALS	Y	Y	2.5
Control	M	74		7.5	NL	N	N	7.1
Control	F	82		5	NL	N	N	7.3
Control	M	55		11.5	NL	N	N	6.8
Control	F	61		20	NL	N	N	7.8
Control	F	59		13	NL	N	N	8.2
Control	M	59		17	NL	N	N	6.9

Table 4.1: Patient demographics of cohort.

Results:

Fluorescence Activated Cell Sorting and RNA Sequencing of Sorted Neuronal Nuclei

The *C9orf72* mutation leads to TDP-43 pathology, which we have shown has large effects on the nuclear transcriptome. To determine whether the *C9orf72* mutation leads to transcriptomic alterations independent of TDP-43 pathology, we isolated neuronal nuclei with intact TDP-43 expression. Frontal neocortex from 7 post-mortem *C9orf72* expansion carriers and 6 neurologically normal controls were used to isolate nuclei for transcriptome-wide analysis. The clinical demographics of these patients can be found in Table 4.1. Nuclei were immunostained for NeuN and TDP-43 and subjected to FAC sorting to isolate NeuN positive, TDP-43 positive nuclei from controls (circled, Fig 4.1A) and *C9orf72* expansion carriers (circled, Fig 4.1B). RNA from equal numbers of TDP-43 positive neuronal nuclei (35,000-100,000) was extracted and amplified to generate barcoded cDNA libraries for 100 or 125bp paired end sequencing on the Illumina HiSeq 2000/2500. Sequences were mapped to the human genome (Gencode GRCh38) using STAR algorithms (version 2.2.4). There were 1.8 billion reads of which 1.065 billion reads mapped uniquely between the 13 libraries, with an average of 90 million uniquely mapped reads per library. These reads were filtered to remove ribosomal and mitochondrial reads and the resulting reads were used for downstream analysis to evaluate differential expression of genes, genic elements and alternative splicing.

Functional transcriptomic alterations associated with the *C9orf72* hexanucleotide repeat expansion

To verify that the variation within the dataset is linked to the mutation status, principal component analysis (PCA) was performed using the read counts of the entire dataset (Fig 4.1C). Principal component #1 (PC1) explained 21% of the variation whereas PC2 explained 19% of the variation (Fig 4.1C). Based on the clustering of the dataset, gender explains the variation in PC1 while the *C9orf72* repeat expansion explains the variation in PC2. Indeed, PC2 strongly correlated with *C9orf72* mRNA expression (Pearson's r=-0.8205; p=0.0006), suggesting that

C9orf72 expression is the major underlying biological variable that contributes to the variation in PC2 (Fig 4.1D). To identify whether there were enriched pathways within the genes that contribute most to PC2, gene ontology analysis was performed using the top 1% of all expressed genes (318 genes) that contribute to PC2. One pathway that is of interest is the cytoplasmic vesicle part or cytoplasmic vesicle membrane (Table 4.2) and is consistent with the functional role of C9orf72 as a vesicle trafficking protein and Rab GTPase (Aoki et al., 2017; D. Zhang et al., 2012).

GO Category	GO Term	No. of Genes	Adj p-value
Molecular Function	11-beta-hydroxysteroid dehydrogenase [NAD(P)] activity	2	2.18E-2
Cellular Component	Cytoplasmic vesicle part	13	4.02E-2
Cellular Component	Cytoplasmic vesicle membrane	12	4.02E-2

Table 4.2: Gene ontology of genes related to top 1% of all genes that contribute to variation in PC2.

Previous studies have found that the G4C2 RNA foci may colocalize with different RNA binding proteins (RBPs), suggesting that sequestration of these RBPs can result in aberrant RNA processing (Conlon et al., 2016; Cooper-Knock et al., 2014; Y. B. Lee et al., 2013). Differential gene expression analysis showed that there are 323 significantly differentially expressed genes with 202 upregulated and 121 downregulated genes linked to the *C9orf72* mutation (Fig 4.1E). Similar to a previous transcriptome analysis on *C9orf72* expansion carriers on the frontal cortex, there were more upregulated genes than downregulated genes within the differentially expressed genes (Prudencio et al., 2015). Notably, the number of differentially expressed genes linked the *C9orf72* repeat expansion was relatively mild compared to the massive transcriptome-wide alterations we previously described associated with TDP-43 pathology in these same cases.

Further annotation shows that these genes are generally involved in synaptic vesicle fusion or vesicle formation (Table 4.3). Additional genes related to vesicle transport and endosomal trafficking are also dysregulated. Interestingly, nine out of these 11 genes are

upregulated. Additional genes that were upregulated were involved in protein aggregation. DNAJB2 is almost exclusively expressed in neurons and has been shown to resolve TDP-43 aggregates by interacting with heat shock protein 70 (H. J. Chen et al., 2016). MGRN1 has been shown to confer cytoprotective effects in an ALS mouse model and can suppress chaperone associated misfolded protein aggregation and toxicity (Chhangani et al., 2016). Given that all *C9orf72* expansion carriers exhibit protein aggregates in the form of dipeptide repeat protein and TDP-43 aggregates, it is possible that these genes are upregulated to help dissociate these aggregates. Lastly, there were also genes that are involved in DNA repair (FAAP20, FAM175A, FANCB, HMGN1) (Ali et al., 2012; Birger et al., 2003; Nomura, Adachi, & Koyama, 2007; B. Wang et al., 2007) and respond to DNA damage (KDM4B and KIN) (Castellini et al., 2017; Miccoli et al., 2003; Young, McDonald, & Hendzel, 2013). Annotation of these genes reveal multiple themes relevant to *C9orf72* including synaptic vesicle formation, endosomal trafficking, chaperone associated protein aggregation, and DNA damage (Table 4.3).

***C9orf72* mutation is associated with mild splicing alterations.**

Previous studies have shown that there are massive alternative splicing changes that are linked to the *C9orf72* mutation (Prudencio et al., 2015). Findings suggesting that RNA binding proteins were colocalizing with RNA foci also support the idea that splicing changes would be rampant in *C9orf72* expansion carriers. To evaluate splicing changes associated with the *C9orf72* mutation, rMATS was used to align junction reads to annotated junctions from GRCh38 assembly. Within our dataset, there were a total of 112 events that were significantly alternatively spliced which included 83 skipped exons, 9 mutually exclusive exons, 2 alternative 3' splice sites and, 8 alternative 5' splice site events and 9 retained introns (Fig 4.1F). These events affected 111 genes, of which only 3 were also significantly differentially expressed. Gene ontology analysis of 111 genes did not result in any significant enriched pathways. Therefore, there were mild splicing changes associated with the *C9orf72* mutation, which is in contrast to previously

described changes. This may be due to technical differences where the use of nuclear RNA in our study likely reduces sensitivity in terms of identifying splicing alterations.

C9orf72 mutation is associated with differentially used 3' UTRs.

RNA foci from *C9orf72* expansion carriers have been shown to sequester different RNA binding proteins (RBP) and suggest the possibility that altered RNA processing is a key feature of *C9orf72* toxicity. RNA binding proteins can bind to different genic elements including exons, introns and untranslated regions (UTR). If RBP activity is altered via titration of these proteins, it is possible the regions that may be bound by RBPs are also differentially expressed. Thus, DEXSeq was used to determine whether the presence of the repeat expansion resulted in differential usage of genic elements. There were a total of 865 significantly differentially used elements linked to the repeat expansion affecting 791 genes. Specifically, there were 610 significantly downregulated elements and 255 significantly upregulated elements (Fig 4.2A). When the elements were annotated as either 5' UTR, exon, intron or 3' UTR, there was a significant enrichment of differentially used 3' UTRs due to the *C9orf72* mutation (13.35% vs 9.78%, χ^2=12.50, p=0.0004) (Fig 4.2B). Given the role of RBPs in binding to 3' UTRs, we hypothesized that the significantly differentially used elements were enriched for RBP binding sites. Thus, hnRNP binding sites, including that of TDP-43 and other hnRNPs were identified using PIPE-CLIP algorithms (Chen B et al. Gen Biol. 2014.) applied to published data sets (Huelga et al., 2012; Tollervey et al., 2011). Indeed, there was an enrichment of differentially used bins that contain TDP-43 and hnRNP binding sites (χ2=6.021, p=0.0141, Fig 4.2C). Thus, the *C9orf72* expansion is associated with differentially used genic elements with RBP binding sites.

Transcriptomic signature due to depletion of TDP-43 pathologic nuclei

Given that neurons with TDP-43 pathology were depleted via FAC sorting prior to RNA-sequencing, we considered the possibility that the transcriptomic alterations associated with the

C9orf72 mutation in this dataset would in part reflect this experimental setup. As shown in Fig 4.3A, *C9orf72* expansion carriers all exhibit TDP-43 pathology, which is reflected as loss of nuclear TDP-43. Thus, carriers have both diseased nuclei (white) and relatively normal nuclei with TDP-43 (blue) wherein we collected normal nuclei and depleted diseased nuclei (red box). On the other hand, non-diseased controls only have normal nuclei with TDP-43 (blue) (Fig 4.3B). As a result, we hypothesized that the depletion of diseased nuclei would be reflected within the transcriptomic alterations associated with the *C9orf72* mutation. Thus, we expected that common genes that were downregulated in *C9orf72* expansion carriers would be upregulated in nuclei without TDP-43 and vice versa. Indeed, among the 118 common significant differentially expressed genes linked to TDP-43 loss and the *C9orf72* mutation, a majority of them followed this pattern (Fig 4.3C). A linear regression analysis showed there was a negative correlation between the gene fold change linked to TDP-43 loss and the gene fold change linked to the *C9orf72* mutation (Pearson's r=-0.5708; $p<0.0001$). Furthermore, this relationship was extended transcriptome-wide where we observed a negative correlation in all commonly expressed genes derived from both data sets (Fig 4.3D). This suggested that depletion of the diseased nuclei in *C9orf72* expansion carriers may explain the transcriptomic alterations observed in *C9orf72* expansion carriers.

Global *C9orf72* associated transcriptome changes are linked to loss of function of the *C9orf72* protein.

Current hypotheses of *C9orf72* toxicity include loss of C9orf72 protein or gain of toxic RNA foci and dipeptide repeat aggregates (S. C. Ling et al., 2013). We have previously shown that the *C9orf72* promoter is hypermethylated in a subset of *C9orf72* expansion carriers and this is neuroprotective in that it is associated with reduced mRNA expression, RNA foci, dipeptide repeat aggregates, improved survival and less cognitive decline in FTD patients (E. Y. Liu et al., 2014; McMillan et al., 2015; Russ et al., 2015), supporting that *C9orf72* toxicity is likely due to a gain of toxic RNA. While we highlighted above specific changes, which appear to be related to

C9orf72, gain of toxic function (genes associated with proteostasis and DNA damage, dysregulation of transcripts with hnRNP 3'UTR binding sites), other changes appeared to be linked to the loss of C9orf72 protein (genes associated with vesicle membranes and endosomal trafficking). However, at a global transcriptional level, it was unclear whether the overall changes associated with the mutation are biased towards effects secondary to a toxic gain of *C9orf72* function or loss of C9orf72 protein.

To determine whether the changes reflect either of these processes, a correlation involving the *C9orf72* promoter methylation and gene expression was done. A model was generated wherein statistical analysis was performed *within C9orf72* expansion carriers versus statistical analysis *between C9orf72* expansion carriers and controls. As an example, using *C9orf72* itself as a gene that contributes to the loss of C9orf72 protein, we predicted that within *C9orf72* expansion carriers, methylation of the *C9orf72* promoter would negatively correlate with *C9orf72* expression (Fig 4.4A). Moreover, we predicted that between group analyses would show that *C9orf72* expression is downregulated compared to controls (Fig 4.4B). This would be an example of a "concordant" gene, which provides evidence that the altered gene expression is linked to the loss of C9orf72 protein. Conversely, using similar logic, if a gene contributes to the gain of toxic function such as the formation of toxic RNA, we would predict that within *C9orf72* expansion carriers, methylation of the *C9orf72* promoter would negatively correlate with toxic RNA (Fig 4.4C). However, between groups, this gene would be upregulated in *C9orf72* expansion carriers compared to controls (Fig 4.4D). This would be an example of a "discordant" gene, which provides evidence that altered expression of this gene is associated with a gain of toxic function.

Using the actual data, a correlation analysis was performed comparing *C9orf72* expression and methylation of the *C9orf72* promoter within our *C9orf72* expansion cohort (Pearson's r=-0.6958; p=0.0825) (Fig 4.4E). Moreover, between-group analysis revealed that *C9orf72* expression is downregulated in *C9orf72* expansion carriers compared to controls as

expected (t-test, p=0.0003) (Fig 4.4F). Another example of a "concordant" gene is synaptophysin (*SYP*), a protein that is involved in synaptic vesicles. The same correlation analysis was done which showed a positive correlation within *C9orf72* expansion carriers (Pearson's r=0.8835; p=0.0083), and an increase in expression in *C9orf72* expansion carriers compared to controls (t-test, p=0.03) (Fig 4.4G-H). Thus, we are able to show that "concordant" genes tend to reflect genes that are related to C9orf72 protein function.

This analysis was extended to all significantly differentially expressed genes where the "within *C9orf72* mutation carrier" methylation correlation value was plotted as a function of "between C9orf72 and control" gene fold change values (Fig 4.4I). Across all significantly differentially expressed genes, there were 267 genes found within the concordant quadrants (blue circles) compared to only 56 genes found in the discordant quadrants (red circles) (Fig 4.4G). The same was found when this analysis was further expanded transcriptome-wide, suggesting that the global transcriptome changes linked to the *C9orf72* mutation are associated with loss of C9orf72 protein (Fig 4.4H). Many "concordant" genes were genes involved in synaptic vesicle formation (i.e. clathrin light chain B, syntaphilin, synaptophysin) and vesicle trafficking (i.e. multivesicular body subunit 12A, RAB40B, member RAS oncogene family). Genes labeled as "discordant" include genes related to inflammatory or apoptotic processes including tec protein tyrosine kinase, tyrosine kinase, and Ring finger protein 152.

Thus two different factors appear to shape this global *C9orf72* transcriptome dataset: (1) the depletion of neurons with TDP-43 pathology prior to RNA-sequencing and (2) the loss of *C9orf72* function. To estimate the relative contributions of these two factors, a multivariate regression analysis was performed wherein significant changes in gene expression associated with the *C9orf72* mutation were related to (1) the changes in gene expression we previously identified due to TDP-43 pathology and (2) the loss of *C9orf72* gene expression due to methylation of the *C9orf72* promoter. Together, these two factors explained the majority of the variance in this dataset (R^2=0.5138; $\beta_{methylation}$=0.838, p<2e-16; $\beta_{TDP\text{-}43}$=-0.499, p<2e-16).

Discussion:

Here we report that in using post-mortem neuronal nuclei without TDP-43 pathology from *C9orf72* expansion carriers, we were able to identify transcriptome changes that link the *C9orf72* mutation with loss of C9orf72 function. We found that the top 1% of genes that contribute most to changes in *C9orf72* expression were involved in cytoplasmic vesicle trafficking. Significant gene expression changes were linked to the *C9orf72* mutation and featured upregulated genes related to synaptic transmission and endosomal/lysosomal trafficking. Indeed, C9orf72 has been shown to be involved in endosomal or lysosomal trafficking and may be a Rab GTPase (Aoki et al., 2017; Farg et al., 2014; D. Zhang et al., 2012). Nine of the 11 genes related to trafficking were upregulated, potentially reflecting a compensatory mechanism to counteract the loss of C9orf72 and loss of this trafficking related protein.

Further analysis correlating methylation and gene expression changes confirmed that the transcriptome changes are linked to loss of *C9orf72* expression. To parse the relative contributions of C9orf72 protein loss or gain of toxic RNA, a methylation correlation between methylation values and expression data was calculated. This correlation analysis indicated that the transcriptome observed in neuronal nuclei from *C9orf72* expansion carriers was driven in large part by the loss of C9orf72 protein as opposed to a toxic gain of function. In contrast, we observed minor splicing alterations and relatively subtle changes in terms of alterations in RNA binding protein function. Thus, the global alterations appear to be most reflective of a loss of C9orf72 function rather than toxic RNA mediated effects.

Genes related to DNA repair either as a response to DNA damage or important to employ DNA repair were also dysregulated. Interestingly, recent papers have shown that DNA damage is activated by the *C9orf72* repeat expansion in both ALS patients and induced pluripotent stem cell derived motor neurons (Farg, Konopka, Ying Soo, Ito, & Atkin, 2017; Lopez-Gonzalez et al., 2016). DNA damage has been a consistent feature among trinucleotide repeat expansion diseases and the dysregulation of DNA repair related genes may reflect a greater role in disease than previously thought (Lopez Castel, Cleary, & Pearson, 2010). Furthermore, alterations in

proteostasis either by failure to misfold proteins properly or failure to degrade proteins contribute to protein aggregation, is a common theme among all neurodegenerative diseases (Douglas & Dillin, 2010; King et al., 2012). In the case of ALS, misfolded TDP-43 aggregates in the cytoplasm, may recruit chaperones including heat shock proteins, and consequently deplete the cell of available chaperones, thereby contributing to cellular dysfunction (Webster, Smith, Shaw, & De Vos, 2017). Indeed, overexpressing heat shock proteins in animal and cell models have prevented TDP-43 aggregation (H. J. Chen et al., 2016; Jackrel et al., 2014; P. Y. Lin et al., 2013), suggesting that altered chaperone activity contributes to neurodegeneration. Thus, although the global transcriptome appears to be driven by the loss of *C9orf72* function, embedded within the *C9orf72* transcriptome are changes in gene expression, which may relate to the toxic functions of *C9orf72.*

In comparing the data from previous transcriptome analyses from *C9orf72* expansion carriers, several differences can be observed. Previous transcriptome analyses have identified pathways involved in inflammation and defense responses among dysregulated genes in the frontal cortices from *C9orf72* expansion carriers (Prudencio et al., 2015). Within our dataset, we did not find enriched inflammatory pathways, which is likely due to the fact that only neurons were being sequenced. It is possible that because *C9orf72* is expressed higher in microglia than neurons, C9orf72 protein function may have a larger effect on microglial populations (O'Rourke et al., 2016). Given the evidence supporting that *C9orf72* is involved inflammatory processes, further cell-specific analyses may prove beneficial to better understanding the effects of mutant *C9orf72* on non-neuronal cells in driving disease pathogenesis (Burberry et al., 2016; J. Jiang et al., 2016; O'Rourke et al., 2016).

Previous studies have found large numbers of splicing changes in *C9orf72* expansion carriers and found motifs within these splicing changes that correspond to hnRNP H binding (Prudencio et al., 2015). However, we did not detect such large changes and instead, found milder splicing changes with motifs that do not correspond to any known RNA foci colocalization partners (data not shown). Current studies debated over the disease relevance of RNA foci as

some animal models overexpressing the repeat induce RNA foci but do not show motor or cognitive defects (O'Rourke et al., 2015; Peters et al., 2015) and models where RNA foci were induced do not die or exhibit noticeable dysregulation (Jain & Vale, 2017; Tran et al., 2015). The lack of splicing changes in our dataset does not support the model wherein neuronal toxicity is linked to the titration of RNA binding proteins within RNA foci.

Overall, our findings argue against RNA toxicity attributed to sequestration of RNA binding proteins. Other disease-related pathways, such as protein aggregation and DNA repair, were observed in this dataset, supporting a potential role for dipeptide repeat protein aggregates and nuclear DNA damage in disease, consistent with experimental studies showing that overexpression of dipeptide repeat proteins are toxic both in cells and in animal models (Lopez-Gonzalez et al., 2016; Mizielinska et al., 2014; Ohki et al., 2017; Wen et al., 2014). While the transcriptome appeared to be shaped in large part by the loss of *C9orf72* function, it is unclear whether the loss of *C9orf72* function contributes to disease and thus understanding whether these changes are functionally linked to disease pathogenesis requires additional experimental studies and validation. Finally, our analysis highlights the complexities associated with molecular studies of post-mortem human tissue wherein issues related to cell identity, experimental design, and admixtures of both gain and loss of function effects can be seen concurrently.

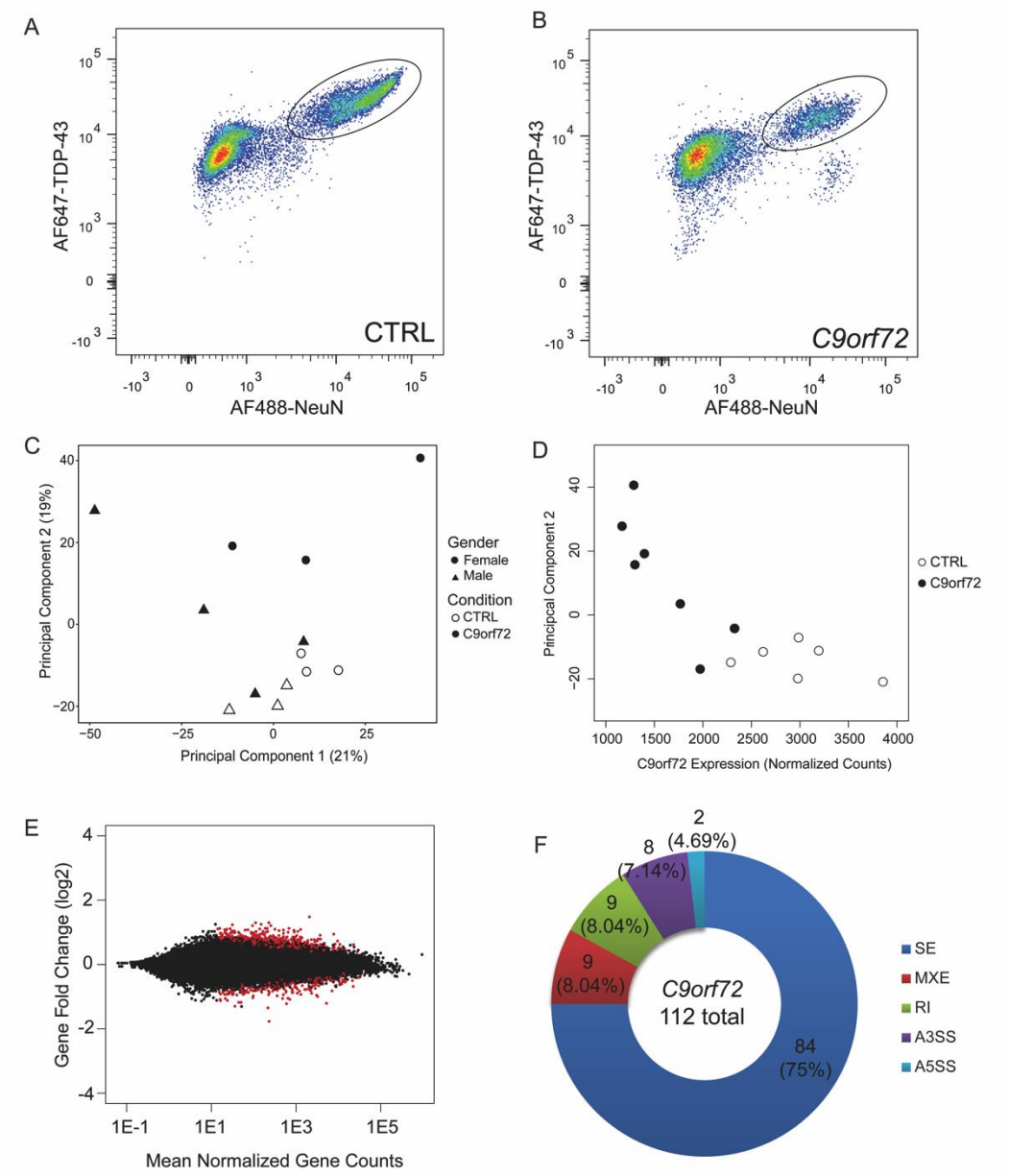

Fig 4.1: ***C9orf72* repeat expansion contributes to mild transcriptome changes**
(A-B) Flow cytometry plots of control patients (A) and *C9orf72* expansion carrier (B) with TDP-43 positive neuronal nuclei circled as collected samples.
(C) Principal component analysis using all expressed genes where shape denotes gender and color denotes condition.
(D) Correlation analysis of principal component 2 and *C9orf72* expression from 7 *C9of72* expansion carriers and 6 controls (Pearson's r = -0.8205; p=0.0006).
(E) MA plot showing gene fold change as a function of mean normalized gene counts with red dots being significantly differentially expressed genes.
(F) Pie chart distribution of all alternative splicing events between *C9orf72* expansion carriers and non-diseased controls.

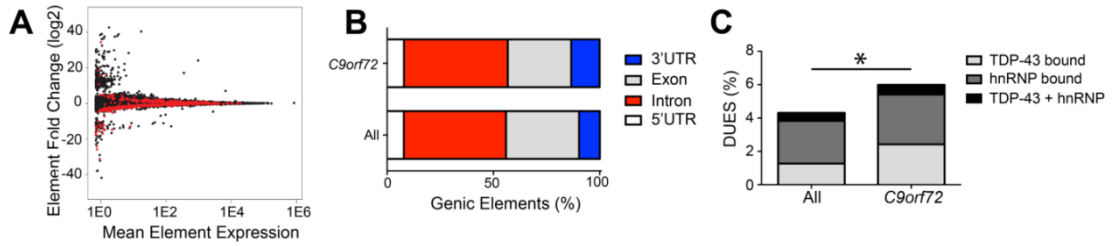

Fig 4.2: ***C9orf72* mutation carriers is associated with differential usage of genic elements with RNA binding protein sites**
(A) MA plot of DEXSeq differentially used genic elements with red being significantly differentially used elements.
(B) Genic elements (5' UTR, exon, intron, 3' UTR) distribution associated with *C9orf72* mutation and all expressed bins.
(C) Differentially used elements that are bound by TDP-43 and hnRNP ($\chi 2=5.617$; p val=0.0178)

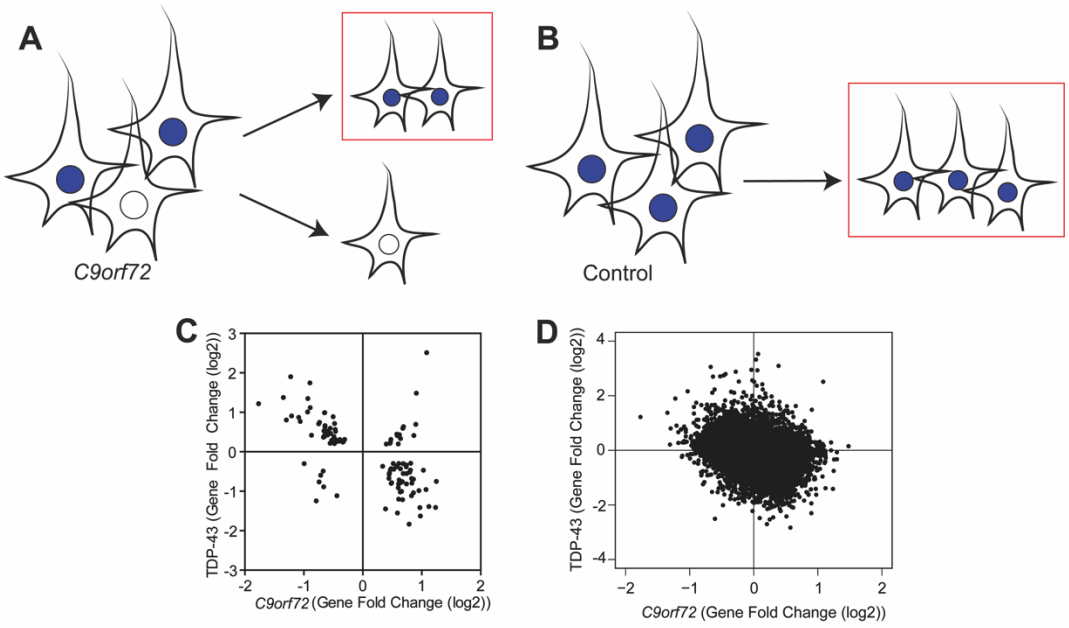

Fig 4.3: ***C9orf72* transcriptome is linked to depletion of TDP-43 pathologic nuclei**
(A) In *C9orf72* expansion carriers, there are nuclei with TDP-43 (blue) and those without TDP-43 (white). For our transcriptome analysis, only nuclei with TDP-43 were collected (red box), thereby depleting nuclei without TDP-43.
(B) In non-diseased controls, all nuclei with TDP-43 were collected (red box).
(C) Gene fold changes of genes that were common significant differentially expressed linked to TDP-43 loss and the *C9orf72* mutation were plotted, showing that genes that were downregulated due to the C9orf72 mutation were upregulated due to TDP-43 loss. Linear regression analysis showed that there was negative correlation (Pearson's r=-0.5708, p<0.0001) between gene fold change linked to TDP-43 loss and the *C9orf72* expansion.
(D) Plotting gene fold changes of expressed genes from both datasets linked to TDP-43 loss and *C9orf72* mutation showed a similar pattern.

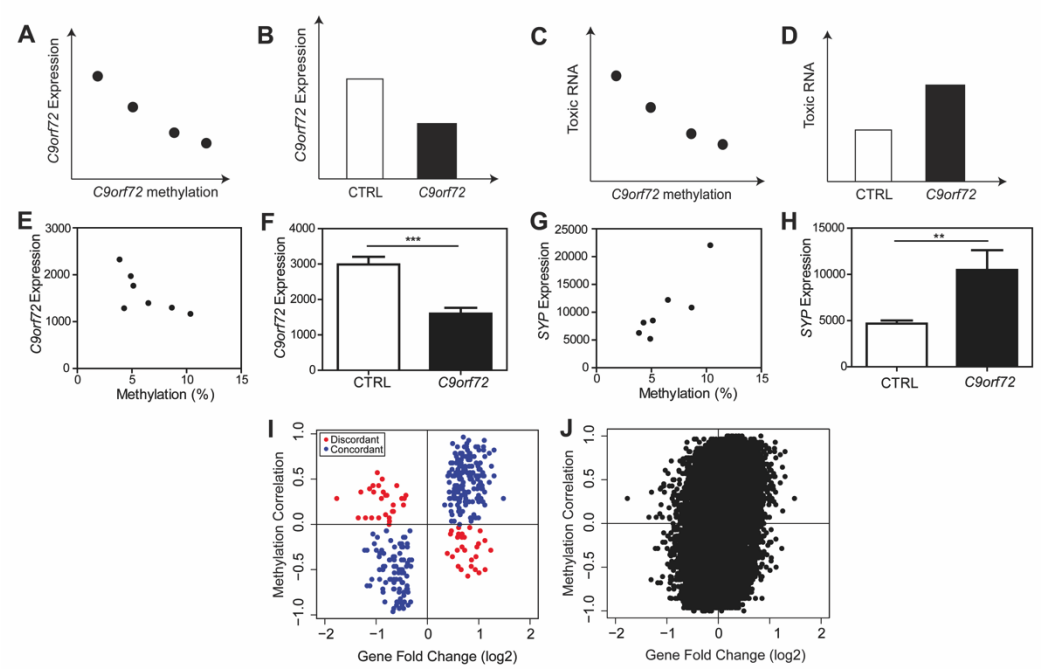

Fig 4.4: ***C9orf72* transcriptome reflects the loss of C9orf72 protein.**
(A) Within *C9orf72* expansion carriers, we predict that *C9orf72* promoter methylation would negatively correlate with *C9orf72* expression.
(B) Between *C9orf72* mutation carriers and non-diseased controls, we predict that *C9orf72* expression is downregulated compared to controls. This would be an example of a “concordant” gene that suggests this gene is linked to loss of C9orf72 protein.
(C) Within *C9orf72* mutation carriers, we predict that *C9orf72* promoter methylation would negative correlate with toxic RNA.
(D) Between *C9orf72* expansion carriers and non-diseased controls, we predict that toxic RNA expression would be upregulated. This would be an example of a “discordant” gene that suggests this gene is linked to gain of toxic RNA.
(E) Correlation analysis between *C9orf72* expression and *C9orf72* promoter methylation within *C9orf72* expansion carriers (Pearson r=-0.6958, p=0.0825) (n=7).
(F) Normalized *C9orf72* expression between *C9orf72* expansion carriers (n=7) and controls (n=6) (t-test, p=0.0003).
(G) Correlation analysis between synaptophysin (*SYP*) expression and *C9orf72* methylation status within *C9orf72* expansion carriers (n=7) (Pearson r=0.8835, p=0.0083).
(H) Normalized *SYP* expression between *C9orf72* expansion carriers (n=7) and controls (n=6) (t-test, p=0.03).
(I) Using 323 significantly differentially expressed genes, the methylation correlation was plotted against gene fold change. Blue dots refer to concordant genes, those reflecting loss of C9orf72 protein and red dots refer to discordant genes, those that reflect gain of toxic RNA.
(J) Genome wide analysis of methylation correlation vs gene fold change using all expressed genes shows the same trend.

Terms of Interest	HGNC symbol	Fold Change (log2)	FDR	Name
SYNAPTIC VESICLE / ENDOCYTOSIS	*BIN1*	0.45	0.01	bridging integrator 1
	DLG4	0.48	0.04	discs large MAGUK scaffold protein 4
	SNPH	0.62	0.01	syntaphilin
	CLTB	0.66	0.03	clathrin light chain B
	SYP	0.94	0.01	synaptophysin
	AHNAK	0.95	0.04	AHNAK nucleoprotein
VESICLE TRANSPORT/ ENDOSOMAL TRAFFICKING	*MVB12A*	0.86	0.05	multivesicular body subunit 12A
	RIN2	0.68	0.01	Ras and Rab interactor 2
	SCAMP2	0.49	0.05	secretory carrier membrane protein 2
	LRPPRC	-0.40	0.00	leurine rich pentratricopeptide repeat containing
	RAB40B	-0.47	0.01	RAB40B, member RAS oncogene family
LYSOSOMAL INVOLVEMENT	*LYST*	-0.31	0.05	lysosomal trafficking regulator
	ASGR1	0.72	0.03	asialoglycoprotein receptor 1
	CLN3	0.76	0.04	CLN3, battenin
ALTERNATIVE SPLICING	*CLK4*	-0.66	0.04	CDC-like kinase 4
	RBM39	-0.40	0.04	RNA binding motif protein 39
	ZRSR2	0.46	0.03	zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2
	SNRPA	0.82	0.03	small nuclear ribonucleoprotein polypeptide A
PROTEIN AGGREGATE	*DNAJB11*	-0.87	0.04	DNAJ (Hsp40) homolog, subfamily B, member 11
	DNAJB2	0.74	0.03	DnaJ heat shock protein family (Hsp40) member B2
	MGRN1	0.74	0.00	mahogunin ring finger 1
DNA REPAIR	*HMGN1*	-0.66	0.04	high mobility group nucleosome binding domain 1
	FANCB	-0.64	0.04	fanconi anemia, complementation group B
	KIN	-0.35	0.04	Kin17 DNA and RNA binding protein
	KDM4B	0.42	0.05	lysine demethylase 4B
	FAAP20	0.68	0.03	Fanconi anemia core complex associated protein 20
	FAM175A	0.73	0.04	family with sequence similarity 175 member A

Table 4.3: Annotation of significantly differentially expressed genes with general terms relevant to *C9orf72* mutation.

Contributions and Acknowledgements:

Elaine Y Liu designed the experiments, developed the Flow-Seq method, performed the RNA-seq experiment, analyzed the RNA-seq data and wrote the paper.

Jenny Russ developed the bioinformatics pipeline analyses.

Edward B Lee designed the experiments, analyzed results and edited the paper.

I would like to thank the University of Pennsylvania Flow Cytometry Core, particularly Paul Hallberg for doing the FAC sorting and the University of Pennsylvania Next Generation Sequencing Core for performing the sequencing. I would like to thank the patients and families who make this work possible.

Chapter 5:

CONCLUSIONS

In my thesis, I have used molecular analyses of neurons from ALS and FTD post-mortem brain to unveil new disease mechanisms that may contribute to disease pathogenesis in the context of the *C9orf72* repeat expansion and loss of nuclear TDP-43. In Chapter 2, I showed the *C9orf72* promoter hypermethylation is present in a subset of *C9orf72* expansion carriers and *C9orf72* promoter hypermethylation is associated with reduced toxic RNA and dipeptide repeat aggregates. This suggests that loss of C9orf72 protein may actually be protective. In Chapter 3, I used molecular phenotyping of neuronal nuclei through FAC sorting from post-mortem ALS/FTD frontal cortex to investigate the role of nuclear TDP-43 loss. Indeed, I have validated many features of TDP-43 biology that, for the first time, has been all shown together in post-mortem human tissue, including massive transcriptome changes, *TARDBP* autoregulation, and selective vulnerability. Importantly, I posit a new disease mechanism whereby loss of nuclear TDP-43 contributes to increased accessibility around repetitive LINE elements and suggest that increased retrotransposon activity may contribute to disease. In Chapter 4, I used the same molecular phenotyping method to show that the *C9orf72* transcriptome is associated with mild gene expression and splicing changes, contrary to a previous study (Prudencio et al., 2015). Furthermore, I show that the *C9orf72* transcriptome is not linked to gain of toxic RNA but rather loss of C9orf72 protein. All in all, my studies suggest that (1) in the context of the *C9orf72* expansion, loss of C9orf72 protein and gain of toxic RNA may not be the main modes of toxicity that contribute to disease and (2) loss of nuclear TDP-43 is linked to massive transcriptome alterations and increased retrotransposon activity that may contribute to disease.

Cell type specificity contributions to neurodegeneration

The adult human brain is a complex organ with a heterogeneous composition consisting of neurons, glial, and endothelial cells. Because each cell type has a different molecular signature, it is important to take this into consideration when one is trying to resolve how different cell types can contribute to neurodegeneration. Indeed, much of the molecular analyses from previous ALS/FTD human patients have used whole brain or laser capture microdissected spinal cord sections (Ladd et al., 2017; Rabin et al., 2010). However, the inflammatory process was identified as one enriched pathway, which is not specific to ALS/FTD and may reflect a secondary consequence of neurodegeneration (Prudencio et al., 2015). Thus, using these methods may not reflect specific disease pathology but an indirect consequence of ongoing gliosis and neuronal degeneration. Although post-mortem brain is very precious and is a translatable model of disease, there are strong caveats that should be considered when using post-mortem brain. Specifically, post-mortem brain is representative of end stage brain, which has cells that have died, cells that are undergoing disease and healthy cells. This calls attention for molecular studies to consider that different cell types within the brain can show varying molecular signatures.

While the focus of the thesis work has been in neurons, other cells types in the brain can contribute to non-cell autonomous disease mechanisms. There is evidence supporting the fact that neurodegenerative diseases including ALS and FTD has a non-cell autonomous disease progression. In the case of superoxide dismutase 1 (*SOD1)* mutations in ALS, which produces misfolded SOD1 protein (Ilieva, Polymenidou, & Cleveland, 2009), mutant SOD1 within primary neurons determines disease onset. However, aberrant synthesis of SOD1 from neighboring cells such as glial cells accelerates disease progression (Boillee et al., 2006; Yamanaka et al., 2008; Zhao et al., 2010). Multiple lines of evidence suggest that loss of *C9orf72* can exhibit a similar disease mechanism. Previous work show that C9orf72 is highly expressed in microglia and loss of C9orf72 in mice does not affect the CNS but rather shows an inflammatory phenotype (Burberry et al., 2016; O'Rourke et al., 2016). It is possible that loss of C9orf72 in microglia can affect disease progression or even promote the onset of disease, similar to SOD1. Thus, future

experiments investigating the effects of C9orf72 loss in microglia co-cultures with either normal or *C9orf72* expansion containing neurons will be important. Specifically, cellular viability or neuronal activity may be important to understand the non-cell autonomous role associated with the *C9orf72* mutation. Furthermore, performing cell type specific transcriptome analyses to reconcile how neighboring cell types can contribute to dying neurons can be an additional layer to understand a potential non-cell autonomous mechanism of disease. This can help unravel novel mechanisms that can explain how other cell types can contribute to neuronal death.

C9orf72 toxicity is likely to be a combination of gain of toxic function and loss of C9orf72 protein.

Since the hexanucleotide repeat expansion in the gene *C9orf72* is the most common genetic cause of sporadic and familial cases of ALS/FTD, there is much debate over the toxic disease mechanism. The observation that *C9orf72* promoter hypermethylation is associated with reduced toxic RNA expression in the form of reduced *C9orf72* associated pathology (RNA foci and dipeptide repeat aggregates), a later age of death and delayed cognitive decline in FTD patients strongly suggest that promoter hypermethylation may be protective in patients (E. Y. Liu et al., 2014; McMillan et al., 2015; Russ et al., 2015). These studies refute the hypothesis that loss of C9orf72 is the main toxic mechanism. However, it is possible that gain of toxic *C9orf72* pathology is more pathogenic than loss of C9orf72, suggesting a fine balance between both disease mechanisms. Importantly, this finding strongly suggests that *C9orf72* hypermethylation can be used as a predictive biomarker for disease and as a stratification method for expansion carriers in clinical trials. Current efforts to develop *C9orf72* associated therapeutics have been focused on antisense oligonucleotides or small molecules that will reduce toxic RNA expression (Donnelly et al., 2013; J. Jiang et al., 2016; Lagier-Tourenne et al., 2013). However, it is possible that in the subset of patients who have hypermethylated *C9orf72* promoters, reduction of toxic *C9orf72* RNA would have no effect, as endogenous promoter hypermethylation naturally silenced

the toxic transcript. *C9orf72* hypermethylation should also be tested in animal or cell models that are used in preclinical studies. Thus, stratifying patient cohorts according to methylation status for clinical trials in which toxic RNA is silenced is paramount. My finding suggests that current efforts to silence toxic *C9orf72* RNA may be a good therapeutic approach and also supports toxic gain of *C9orf72* associated pathology as the main potential disease mechanism.

C9orf72 promoter hypermethylation: Cause and Effect?

Interestingly, roughly 1/3 of tested *C9orf72* expansion carriers within our patient cohort were hypermethylated but the mechanism mediating *C9orf72* methylation in a subset of patients is currently unclear. It is possible that there is a quantitative trait loci (QTL) upstream of the promoter region that is associated with promoter hypermethylation. Using a large cohort of ALS/FTD patients with variably methylated promoter regions for sequencing to identify either a QTL or expression QTL would help us understand what factors contribute to *C9orf72* promoter methylation.

My study shows the correlative effects of methylation but the causative role has not yet been investigated. Further experiments manipulating DNA methylation at the *C9orf72* locus would be ideal to understand the effects of C9orf72 methylation, namely, whether methylation causes *C9orf72* pathology reduction. It is possible that other epigenetic factors including histone marks identified in the *C9orf72* promoter (Belzil, Bauer, et al., 2013) may also reduce *C9orf72* pathology. However, because methylation tends to be a more stable silencing mechanism, it is more likely that methylation is the causative factor in reducing toxic mRNA expression and *C9orf72* pathology. From a therapeutic standpoint, stable methylation of *C9orf72* should be considered. Because localized methylation techniques have not been very stable (Vojta et al., 2016), it is possible that prior manipulation of the histone modifiers may promote prolonged methylation status around the *C9orf72* locus. Thus, more studies to investigate the role of *C9orf72* promoter hypermethylation should be considered.

Additionally, a well-known fact about *C9orf72* transcription is that the repeat expansion can also be transcribed in the antisense direction forming antisense RNA foci and DPR peptides. My study has not yet investigated the effects on either of these factors but it is possible that *C9orf72* promoter hypermethylation upstream of the repeat in the sense direction may not influence transcription from the antisense direction, thereby resulting in antisense *C9orf72* pathology. While the role of antisense RNA foci is still unclear, the formation of antisense DPR peptides have been shown to be pathogenic in cellular models. Indeed, arginine rich DPRs including proline-arginine, produced from the antisense direction (Mizielinska et al., 2014; Wen et al., 2014) were toxic. Further studies are necessary to understand how *C9orf72* hypermethylation occurs and how it affects the cell.

C9orf72 protein loss is a cellular stressor that may make neurons vulnerable to TDP-43 pathology.

Previous transcriptomic studies using whole brain from sporadic ALS and FTD patients bearing the *C9orf72* expansion show global splicing changes and alternative polyadenylation sites, all of which were signs of RNA binding protein dysregulation (Prudencio et al., 2015). However, I did not see aberrant splicing activity linked the *C9orf72* mutation. This raises the possibility that RNA binding protein titration hypothesis does not contribute to transcriptome alterations but rather suggests that gain of toxic *C9orf72* dipeptide repeat pathology may be responsible for *C9orf72* toxicity. Given that I did not observe many transcriptome changes, it is possible that dipeptide repeat pathology may be the contributing disease mechanism. Indeed, a previous report shows that RNA foci do not contribute to toxicity and it may be the presence of dipeptide repeat aggregates that are more toxic (Tran et al., 2015). Overexpression of dipeptide repeat aggregates in both animal and cell models confer toxicity and contribute to nucleolar stress and nucleocytoplasmic transport defects (Freibaum et al., 2015; Jovicic et al., 2015; May et al., 2014; Tao et al., 2015; Wen et al., 2014) or may even impair nuclear transport of TDP-43

(Khosravi et al., 2017). However, there is conflicting evidence regarding the toxic role of DPRs because DPR aggregates do not correlate with degenerating regions of the brain (Mackenzie et al., 2013; Mackenzie et al., 2015). DPR aggregates may act similarly to amyloid beta plaques in AD (Edbauer & Haass, 2016). Both pathologies do not correlate with degenerating areas but this lack of correlation can be explained by proposing that (1) soluble forms of DPRs are toxic or (2) aggregates all over the brain reflect the long-term progression of DPR deposition during the disease course. According to the GENFI imaging cohort where ALS and FTD patients carrying progranulin, tau and *C9orf72* mutations were imaged longitudinally, most tau or progranulin mutation carriers exhibit brain atrophy within 5-10 years before onset compared to non-carriers. However, *C9orf72* expansion carriers show brain atrophy in multiple areas 25 years before disease onset. It is possible that DPR pathology may already be occurring in the brain much earlier than disease diagnosis (Rohrer et al., 2015). Thus, I cannot rule out that DPRs may still contribute to disease toxicity.

This raises the possibility that the *C9orf72* mutation is merely a single hit in a multi-hit scenario that contributes to ALS/FTD. Perhaps the *C9orf72* mutation contributes to toxicity through (1) dipeptide repeat aggregates compromising nucleocytoplasmic transport resulting in nucleolar stress, (2) loss of *C9orf72* protein that preferentially affects microglia that may affect neuronal loss, and (3) oxidative stress from aging. With multiple hits, this can compromise the cell and somehow make it more susceptible to TDP-43 pathology, which results in the eventual demise of the neuron. There is evidence suggesting that C9orf72 repeat pathology predates TDP-43 pathology in patients (Proudfoot et al., 2014). Because TDP-43 pathology is the best correlate to neurodegeneration in ALS/FTD, it is likely that TDP-43 pathology is the toxic killer of neurons. However, further experimentation testing this hypothesis would be difficult considering that the mechanism whereby TDP-43 pathology occurs is unclear.

Cell type specific analysis to study TDP-43 loss shows global transcriptome crisis.

In my thesis, I was able to validate the role of nuclear TDP-43 in post-mortem human brain. I showed that loss of TDP-43 contributes to transcriptomic crisis, an observation that was observed in many animal and cell models but not yet in post-mortem brain. TDP-43 is involved in suppression of repetitive elements (W. Li et al., 2012). I show that loss of nuclear TDP-43 was associated with open chromatin around repetitive elements and that aberrant expression of TDP-43 was able to reactivate LINE retrotransposition. These findings are quite significant to the field for several reasons: (1) I show that cell type specific analyses on post-mortem human tissue was able to identify molecular changes associated with TDP-43 proteinopathies, (2) validate TDP-43 function in post-mortem brain, and (3) exhibits a fundamental phenomenon that may be a consequence of cellular stress.

Nuclei sorting from post-mortem brain can be implemented to answer other biological questions related to nuclear changes in disease. Given that chromatin changes, genomic instability, and RNA differences are common themes among neurodegeneration, there is a need to use cell-type specific analyses to determine how and why some cell types are more vulnerable than others. In the context of ALS/FTD research, this study assessed molecular differences linked to TDP-43 loss in post-mortem brain and provides a resource that can be used by other investigators to determine what changes observed in animal or cell disease models are relevant to humans. Thus, the method I developed to study nuclear changes in post-mortem human nuclei can be easily implemented for other cell-type specific studies and the results I produced can also be used as a general resource for ALS/FTD researchers studying TDP-43.

L1 retrotransposition in TDP-43 pathologic nuclei: implications?

The role of L1 retrotransposition in human brain has been investigated but how TDP-43 relates to increased L1 retrotransposition is unclear. L1 retrotransposition has been shown to occur in embryonic stem cells such that germline L1 insertions can occur but recent studies have expanded to show this occurs in somatic tissue such as human brain (Muotri et al., 2005).

Indeed, multiple studies have used bulk sequencing or single cell sequencing to identify L1 insertion events within hippocampus and caudate nucleus of post-mortem brain and unequivocally shown that L1 insertion happens but the rate of insertion is unclear (Baillie et al., 2011; Evrony, Lee, Park, & Walsh, 2016; Upton et al., 2015).

Based on my results, it is possible that loss of TDP-43 can be a cellular stressor that leads to increased retrotransposition within a subset of neuronal nuclei. What is currently unclear regarding the relevance of L1 retrotransposition events in TDP-43 pathologic nuclei is: (1) where are the potential L1 insertions in pathologic nuclei and (2) what are the consequences of these L1 expression or insertion events in the genome of TDP-43 pathologic nuclei? L1 insertion events typically lead to truncated L1 forms, which can impact RNA expression of neighboring genes. Oftentimes, there is a 5' truncation of the inserted L1 element (Hancks & Kazazian, 2016). Firstly, L1 insertion within a specific gene can result in a gene product that incorporates that L1 polyA tail, leading to a truncated mRNA that can be subject to nonsense mediated decay (NMD). This will inactivate that gene's function. It is also possible that insertion of the L1 element leads to altered gene splicing, which can also change the mRNA transcript and potential protein product if it is not subject to NMD. Indeed, L1 insertion events within neurons have been found in genes highly expressed within the brain, suggesting that these insertion events can affect neuronal function (Richardson, Morell, & Faulkner, 2014). Secondly, L1 insertion events can create new promoter or enhancer regions within the inserted gene that can promote aberrant transcription. Thus, it is important to understand how potential L1 insertion events can alter RNA expression surrounding these genes in the context of TDP-43 pathology. In order to do this, DNA and RNA can be extracted from pathologic neuronal nuclei and normal nuclei to perform single cell RNA-seq and retrotransposon capture-seq. In doing so, one can identify the genes where L1 insertion events and determine whether the insertion events altered RNA expression from neighboring genes. These target genes can be tested in *in vitro* models of disease to determine disease relevance. However, it is possible that L1 insertion is not the only method whereby increased L1 accessibility can contribute to disease. An alternate hypothesis as to how increased L1 openness

can contribute to disease is the mere fact that expressing the L1 element results in the translation of ORF2, which has endonuclease activity. It is possible that because of this activity, there could be random DNA breaks that are created throughout the genome even without genomic integration. Thus, it is possible that toxicity can occur as a result of genomic integration that can interrupt RNA processing or ORF2 activity that can start creating DNA breaks. Nonetheless, both may contribute to disease pathogenesis.

My results in this study can also be interpreted such that LINE reactivation is simply a fundamental phenomenon of cellular stress rather than a direct effect of TDP-43. Aging and DNA damage has been frequently linked to heterochromatin loss and reactivation of transposable elements (De Cecco et al., 2013; Dillon, 2004; K. T. Smith & Workman, 2012; Van Meter et al., 2014; Wood et al., 2016). In Alzheimer's disease, hyperphosphorylated tau is associated with loss of heterochromatin, oxidative damage and chromatin relaxation (Frost et al., 2014; Mansuroglu et al., 2016). In ALS/FTD patients bearing the *C9orf72* mutation, there was increased expression of repetitive elements (Prudencio et al., 2017). Thus, it appears that age and cellular stress can contribute to heterochromatin loss and reactivation of transposable elements. Further experimentation using cellular stressors including hydrogen peroxide, sodium arsenite or sorbitol can help determine whether cellular stress can reactivate LINES. When compounded with disease-associated stresses such as DNA damage associated with *C9orf72* or TDP-43 disease pathology, it would make sense that chromatin relaxation around formerly heterochromatinized regions including repeat elements would occur.

Because we observed that there is increased LINE1 DNA content in post-mortem brain without TDP-43 and aberrant expression of TDP-43 causes increased retrotransposition, it is possible that increased retrotransposition is a signature of all TDP-43 proteinopathies. In using the single cell RNA-seq and RC-seq in diseased patients, it is unlikely that the same L1 insertion events can be identified between patients. Thus, it would be difficult to identify consensus disease associated genes affected by L1 insertion events. In this case, it is possible to target the event upstream of retroviral insertion by using anti-retroviral therapies (ART) to reverse retrotransposon

activity. What is currently unclear is whether retroviral activation causes or is a consequence of TDP-43 pathology. My cellular model where I knocked out TDP-43 suggests that loss of TDP-43 leads to increased retrotransposition but I cannot rule out that the cellular stress associated with TDP-43 loss leads to retrotransposition as well. Given that many aging and cellular stressing related processes are associated with heterochromatin loss and increased repetitive element expression, TDP-43 pathology can merely exacerbate retrotransposition but does not cause it. Thus, one can imagine that ART given in prodromal ALS populations can prevent increased retrotransposition from happening later on. Indeed, there is a clinical trial testing ART in sporadic ALS patients to determine if there are any beneficial effects. Considering that there currently is no cure for ALS/FTD and only palliative care at the moment, there is a push for therapeutic advances for these patients.

In summary, given that there are few therapies for ALS and FTD patients, there is a need to develop therapeutics or biomarkers that will be used in the clinic to alleviate the detrimental disease effects in these patients. The studies that were performed here revolve around two important genetic and pathologic contributing factors to disease: the *C9orf72* repeat expansion and the loss of nuclear TDP-43. From these studies, a transcriptomic signature related to TDP-43 function and the *C9orf72* expansion was highlighted and can be available as a general resource for all ALS/FTD researchers who want to compare animal and cell based models of TDP-43 proteinopathies to post-mortem human tissue analyses. Hopefully, these analyses can be used to discover novel disease mechanisms, potential therapeutics, and biomarkers that can be extended into the clinic.

APPENDIX:

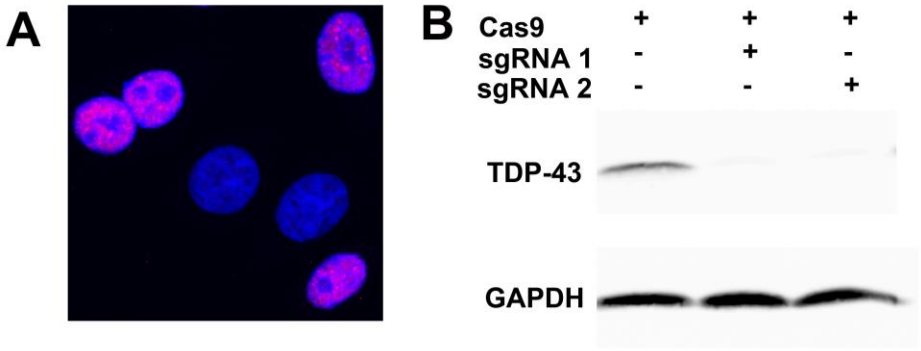

Fig A1: **TDP-43 knock-out in HeLa cells**
(A) Immunocytochemistry of CRISPR/Cas9 engineered cells targeting *TARDBP* where TDP-43 (red) and DAPI (blue) was stained showing that TDP-43 is completely knocked out in a subset of cells. (B) Western blot of cells that have been engineered with CRISPR/Cas9 to knock-out TDP-43 and overexpressing pEF06R 9 days after transfection when retrotransposition activity is assessed.

BIBLIOGRAPHY

Al-Mahdawi, S., Pinto, R. M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., . . . Pook, M. (2008). The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum Mol Genet, 17*(5), 735-746. doi:10.1093/hmg/ddm346

Alami, N. H., Smith, R. B., Carrasco, M. A., Williams, L. A., Winborn, C. S., Han, S. S., . . . Taylor, J. P. (2014). Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron, 81*(3), 536-543. doi:10.1016/j.neuron.2013.12.018

Alfahad, T., & Nath, A. (2013). Retroviruses and amyotrophic lateral sclerosis. *Antiviral Res, 99*(2), 180-187. doi:10.1016/j.antiviral.2013.05.006

Ali, A. M., Pradhan, A., Singh, T. R., Du, C., Li, J., Wahengbam, K., . . . Meetei, A. R. (2012). FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway. *Blood, 119*(14), 3285-3294. doi:10.1182/blood-2011-10-385963

Almeida, S., Gascon, E., Tran, H., Chou, H. J., Gendron, T. F., Degroot, S., . . . Gao, F. B. (2013). Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol, 126*(3), 385-399. doi:10.1007/s00401-013-1149-y

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., & Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol, 18*(12), 1435-1440. doi:10.1038/nsmb.2143

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature genetics, 23*(2), 185-188. doi:10.1038/13810

Amlie-Wolf, A., Ryvkin, P., Tong, R., Dragomir, I., Suh, E., Xu, Y., . . . Lee, E. B. (2015). Transcriptomic Changes Due to Cytoplasmic TDP-43 Expression Reveal Dysregulation of Histone Transcripts and Nuclear Chromatin. *PLoS One, 10*(10), e0141836. doi:10.1371/journal.pone.0141836

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res, 22*(10), 2008-2017. doi:10.1101/gr.133744.111

Anderson, P., & Ivanov, P. (2014). tRNA fragments in human health and disease. *FEBS Lett, 588*(23), 4297-4304. doi:10.1016/j.febslet.2014.09.001

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. .

Aoki, Y., Manzano, R., Lee, Y., Dafinca, R., Aoki, M., Douglas, A. G. L., . . . Wood, M. J. A. (2017). C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. *Brain : a journal of neurology, 140*(4), 887-897. doi:10.1093/brain/awx024

Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., . . . Petrucelli, L. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron, 77*(4), 639-646. doi:10.1016/j.neuron.2013.02.004

Avendano-Vazquez, S. E., Dhir, A., Bembich, S., Buratti, E., Proudfoot, N., & Baralle, F. E. (2012). Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev, 26*(15), 1679-1684. doi:10.1101/gad.194829.112

Ayala, Y. M., De Conti, L., Avendano-Vazquez, S. E., Dhir, A., Romano, M., D'Ambrogio, A., . . . Baralle, F. E. (2011). TDP-43 regulates its mRNA levels through a negative feedback loop. *The EMBO journal, 30*(2), 277-288. doi:10.1038/emboj.2010.310

Ayyanathan, K., Lechner, M. S., Bell, P., Maul, G. G., Schultz, D. C., Yamada, Y., . . . Rauscher, F. J., 3rd. (2003). Regulated recruitment of HP1 to a euchromatic gene induces

mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev, 17*(15), 1855-1869. doi:10.1101/gad.1102803

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., . . . Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol, 513*(5), 532-541. doi:10.1002/cne.21974

Babushok, D. V., & Kazazian, H. H., Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat, 28*(6), 527-539. doi:10.1002/humu.20486

Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., . . . Faulkner, G. J. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature, 479*(7374), 534-537. doi:10.1038/nature10531

Barmada, S. J., Serio, A., Arjun, A., Bilican, B., Daub, A., Ando, D. M., . . . Finkbeiner, S. (2014). Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat Chem Biol, 10*(8), 677-685. doi:10.1038/nchembio.1563

Barmada, S. J., Skibinski, G., Korb, E., Rao, E. J., Wu, J. Y., & Finkbeiner, S. (2010). Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J Neurosci, 30*(2), 639-649. doi:10.1523/JNEUROSCI.4988-09.2010

Barres, B. A. (2014). Designing and troubleshooting immunopanning protocols for purifying neural cells. *Cold Spring Harb Protoc, 2014*(12), 1342-1347. doi:10.1101/pdb.ip073999

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D. J., . . . Swanson, M. S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol Cell, 56*(2), 311-322. doi:10.1016/j.molcel.2014.08.027

Batra, R. H., K.; Vu, A.;Rabin, S.J.; Baughn, M.W.; Libby, R.T.; Hoon, S.; Ravits, J.; Yeo, G.W. (2016). Gene Expression Signatures of Sporadic ALS Motor Neuron Populations. *bioRxiv*. doi:https://doi.org/10.1101/038448

Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., . . . Moran, J. V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell, 141*(7), 1159-1170. doi:10.1016/j.cell.2010.05.021

Belzil, V. V., Bauer, P. O., Prudencio, M., Gendron, T. F., Stetler, C. T., Yan, I. K., . . . Petrucelli, L. (2013). Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol, 126*(6), 895-905. doi:10.1007/s00401-013-1199-1

Belzil, V. V., Gendron, T. F., & Petrucelli, L. (2013). RNA-mediated toxicity in neurodegenerative disease. *Mol Cell Neurosci, 56*, 406-419. doi:10.1016/j.mcn.2012.12.006

Bentmann, E., Haass, C., & Dormann, D. (2013). Stress granules in neurodegeneration--lessons learnt from TAR DNA binding protein of 43 kDa and fused in sarcoma. *FEBS J, 280*(18), 4348-4370. doi:10.1111/febs.12287

Benussi, L., Rossi, G., Glionna, M., Tonoli, E., Piccoli, E., Fostinelli, S., . . . Ghidoni, R. (2013). C9ORF72 Hexanucleotide Repeat Number in Frontotemporal Lobar Degeneration: A Genotype-Phenotype Correlation Study. *Journal of Alzheimer's disease : JAD*. doi:10.3233/JAD-131028

Bernard, A., Lubbers, L. S., Tanis, K. Q., Luo, R., Podtelezhnikov, A. A., Finney, E. M., . . . Lein, E. S. (2012). Transcriptional architecture of the primate neocortex. *Neuron, 73*(6), 1083-1099. doi:10.1016/j.neuron.2012.03.002

Birger, Y., West, K. L., Postnikov, Y. V., Lim, J. H., Furusawa, T., Wagner, J. P., . . . Bustin, M. (2003). Chromosomal protein HMGN1 enhances the rate of DNA repair in chromatin. *The EMBO journal, 22*(7), 1665-1675. doi:10.1093/emboj/cdg142

Boeve, B. F., Boylan, K. B., Graff-Radford, N. R., DeJesus-Hernandez, M., Knopman, D. S., Pedraza, O., . . . Rademakers, R. (2012). Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. *Brain : a journal of neurology, 135*(Pt 3), 765-783. doi:10.1093/brain/aws004

Boillee, S., Yamanaka, K., Lobsiger, C. S., Copeland, N. G., Jenkins, N. A., Kassiotis, G., . . . Cleveland, D. W. (2006). Onset and progression in inherited ALS determined by motor neurons and microglia. *Science, 312*(5778), 1389-1392. doi:10.1126/science.1123511

Bowen, L. N., Tyagi, R., Li, W., Alfahad, T., Smith, B., Wright, M., . . . Nath, A. (2016). HIV-associated motor neuron disease: HERV-K activation and response to antiretroviral therapy. *Neurology, 87*(17), 1756-1762. doi:10.1212/WNL.0000000000003258

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol, 112*(4), 389-404. doi:10.1007/s00401-006-0127-z

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., . . . et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell, 69*(2), 385.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods, 10*(12), 1213-1218. doi:10.1038/nmeth.2688

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., . . . Iwamoto, K. (2014). Increased l1 retrotransposition in the neuronal genome in schizophrenia. *Neuron, 81*(2), 306-313. doi:10.1016/j.neuron.2013.10.053

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., & Baralle, F. E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *The EMBO journal, 20*(7), 1774-1784. doi:10.1093/emboj/20.7.1774

Burberry, A., Suzuki, N., Wang, J. Y., Moccia, R., Mordes, D. A., Stewart, M. H., . . . Eggan, K. (2016). Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med, 8*(347), 347ra393. doi:10.1126/scitranslmed.aaf6038

Burda, J. E., & Sofroniew, M. V. (2014). Reactive gliosis and the multicellular response to CNS damage and disease. *Neuron, 81*(2), 229-248. doi:10.1016/j.neuron.2013.12.034

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., . . . Pandolfo, M. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science, 271*(5254), 1423-1427.

Cascella, R., Capitini, C., Fani, G., Dobson, C. M., Cecchi, C., & Chiti, F. (2016). Quantification of the Relative Contributions of Loss-of-function and Gain-of-function Mechanisms in TAR DNA-binding Protein 43 (TDP-43) Proteinopathies. *The Journal of biological chemistry, 291*(37), 19437-19448. doi:10.1074/jbc.M116.737726

Castaldo, I., Pinelli, M., Monticelli, A., Acquaviva, F., Giacchetti, M., Filla, A., . . . Cocozza, S. (2008). DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *Journal of medical genetics, 45*(12), 808-812. doi:10.1136/jmg.2008.058594

Castellini, L., Moon, E. J., Razorenova, O. V., Krieg, A. J., von Eyben, R., & Giaccia, A. J. (2017). KDM4B/JMJD2B is a p53 target gene that modulates the amplitude of p53 response after DNA damage. *Nucleic acids research, 45*(7), 3674-3692. doi:10.1093/nar/gkw1281

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., . . . Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev, 28*(13), 1397-1409. doi:10.1101/gad.241661.114

Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X., Wong, S. T., Qin, J., & Zoghbi, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science, 320*(5880), 1224-1229. doi:10.1126/science.1153252

Chen, B., Yun, J., Kim, M. S., Mendell, J. T., & Xie, Y. (2014). PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol, 15*(1), R18. doi:10.1186/gb-2014-15-1-r18

Chen, H. J., Mitchell, J. C., Novoselov, S., Miller, J., Nishimura, A. L., Scotter, E. L., . . . Shaw, C. E. (2016). The heat shock response plays an important role in TDP-43 clearance:

evidence for dysfunction in amyotrophic lateral sclerosis. *Brain : a journal of neurology, 139*(Pt 5), 1417-1432. doi:10.1093/brain/aww028

Chhangani, D., Endo, F., Amanullah, A., Upadhyay, A., Watanabe, S., Mishra, R., . . . Mishra, A. (2016). Mahogunin ring finger 1 confers cytoprotection against mutant SOD1 aggresomes and is defective in an ALS mouse model. *Neurobiol Dis, 86*, 16-28. doi:10.1016/j.nbd.2015.11.017

Chio, A., Borghero, G., Restagno, G., Mora, G., Drepper, C., Traynor, B. J., . . . Sabatelli, M. (2012). Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. *Brain : a journal of neurology, 135*(Pt 3), 784-793. doi:10.1093/brain/awr366

Chiu, I. M., Barrett, L. B., Williams, E. K., Strochlic, D. E., Lee, S., Weyer, A. D., . . . Woolf, C. J. (2014). Transcriptional profiling at whole population and single cell levels reveals somatosensory neuron molecular diversity. *Elife, 3*. doi:10.7554/eLife.04660

Chung, C. Y., Seo, H., Sonntag, K. C., Brooks, A., Lin, L., & Isacson, O. (2005). Cell type-specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. *Hum Mol Genet, 14*(13), 1709-1725. doi:10.1093/hmg/ddi178

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., & Kabashi, E. (2013a). Loss of function of C9orf72 causes motor deficits in a zebrafish model of amyotrophic lateral sclerosis. *Annals of neurology, 74*(2), 180-187. doi:10.1002/ana.23946

Ciura, S., Lattante, S., Le Ber, I., Latouche, M., Tostivint, H., Brice, A., & Kabashi, E. (2013b). Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. *Annals of neurology*. doi:10.1002/ana.23946

Cleary, J. D., & Ranum, L. P. (2013). Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum Mol Genet, 22*(R1), R45-51. doi:10.1093/hmg/ddt371

Clouaire, T., & Stancheva, I. (2008). Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol Life Sci, 65*(10), 1509-1522. doi:10.1007/s00018-008-7324-y

Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W. Y., Gerhardt, J., . . . Jaffrey, S. R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science, 343*(6174), 1002-1005. doi:10.1126/science.1245831

Collins, S. C., Bray, S. M., Suhl, J. A., Cutler, D. J., Coffee, B., Zwick, M. E., & Warren, S. T. (2010). Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *American journal of medical genetics. Part A, 152A*(10), 2512-2520. doi:10.1002/ajmg.a.33626

Collins, S. C., Coffee, B., Benke, P. J., Berry-Kravis, E., Gilbert, F., Oostra, B., . . . Warren, S. T. (2010). Array-based FMR1 sequencing and deletion analysis in patients with a fragile X syndrome-like phenotype. *PLoS One, 5*(3), e9476. doi:10.1371/journal.pone.0009476

Conlon, E. G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N. A., & Manley, J. L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife, 5*. doi:10.7554/eLife.17820

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74. doi:10.1038/nature11247

Cooper-Knock, J., Bury, J. J., Heath, P. R., Wyles, M., Higginbottom, A., Gelsthorpe, C., . . . Shaw, P. J. (2015). C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. *PLoS One, 10*(5), e0127376. doi:10.1371/journal.pone.0127376

Cooper-Knock, J., Hewitt, C., Highley, J. R., Brockington, A., Milano, A., Man, S., . . . Shaw, P. J. (2012). Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain : a journal of neurology, 135*(Pt 3), 751-764. doi:10.1093/brain/awr365

Cooper-Knock, J., Higginbottom, A., Connor-Robson, N., Bayatti, N., Bury, J. J., Kirby, J., . . . Shaw, P. J. (2013). C9ORF72 transcription in a frontotemporal dementia case with two expanded alleles. *Neurology, 81*(19), 1719-1721. doi:10.1212/01.wnl.0000435295.41974.2e

Cooper-Knock, J., Kirby, J., Ferraiuolo, L., Heath, P. R., Rattray, M., & Shaw, P. J. (2012). Gene expression profiling in human neurodegenerative disease. *Nat Rev Neurol, 8*(9), 518-530. doi:10.1038/nrneurol.2012.156

Cooper-Knock, J., Walsh, M. J., Higginbottom, A., Robin Highley, J., Dickman, M. J., Edbauer, D., . . . Shaw, P. J. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain : a journal of neurology, 137*(Pt 7), 2040-2051. doi:10.1093/brain/awu120

Cosenza, M. A., Zhao, M. L., Si, Q., & Lee, S. C. (2002). Human brain parenchymal microglia express CD14 and CD45 and are productively infected by HIV-1 in HIV-1 encephalitis. *Brain Pathol, 12*(4), 442-455.

Costessi, L., Porro, F., Iaconcig, A., & Muro, A. F. (2014). TDP-43 regulates beta-adducin (Add2) transcript stability. *RNA Biol, 11*(10), 1280-1290. doi:10.1080/15476286.2014.996081

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Marchetto, M. C., Muotri, A. R., Mu, Y., . . . Gage, F. H. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci U S A, 108*(51), 20382-20387. doi:10.1073/pnas.1100273108

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., . . . Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature, 460*(7259), 1127-1131. doi:10.1038/nature08248

Courchaine, E. M., Lu, A., & Neugebauer, K. M. (2016). Droplet organelles? *The EMBO journal, 35*(15), 1603-1612. doi:10.15252/embj.201593517

Coyle, J. T., & Puttfarcken, P. (1993). Oxidative stress, glutamate, and neurodegenerative disorders. *Science, 262*(5134), 689-695.

Curran, S., McKay, J. A., McLeod, H. L., & Murray, G. I. (2000). Laser capture microscopy. *Mol Pathol, 53*(2), 64-68.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., . . . Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A, 112*(23), 7285-7290. doi:10.1073/pnas.1507125112

Datta, S., Malhotra, L., Dickerson, R., Chaffee, S., Sen, C. K., & Roy, S. (2015). Laser capture microdissection: Big data from small samples. *Histol Histopathol, 30*(11), 1255-1269. doi:10.14670/HH-11-622

Daughters, R. S., Tuttle, D. L., Gao, W., Ikeda, Y., Moseley, M. L., Ebner, T. J., . . . Ranum, L. P. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. *PLoS Genet, 5*(8), e1000600. doi:10.1371/journal.pgen.1000600

Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., & Housman, D. E. (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A, 94*(14), 7388-7393.

De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., . . . Sedivy, J. M. (2013). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell, 12*(2), 247-256. doi:10.1111/acel.12047

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., . . . Bennett, D. A. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci, 17*(9), 1156-1163. doi:10.1038/nn.3786

de Vries, B. B., Jansen, C. C., Duits, A. A., Verheij, C., Willemsen, R., van Hemel, J. O., . . . Halley, D. J. (1996). Variable FMR1 gene methylation of large expansions leads to variable phenotype in three males from one fragile X family. *Journal of medical genetics, 33*(12), 1007-1010.

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev, 25*(10), 1010-1022. doi:10.1101/gad.2037511

DeJesus-Hernandez, M., Finch, N. A., Wang, X., Gendron, T. F., Bieniek, K. F., Heckman, M. G., . . . Rademakers, R. (2017). In-depth clinico-pathological examination of RNA foci in a large cohort of C9ORF72 expansion carriers. *Acta Neuropathol, 134*(2), 255-269. doi:10.1007/s00401-017-1725-7

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., . . . Rademakers, R. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron, 72*(2), 245-256. doi:10.1016/j.neuron.2011.09.011

Devys, D., Biancalana, V., Rousseau, F., Boue, J., Mandel, J. L., & Oberle, I. (1992). Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. *American journal of medical genetics, 43*(1-2), 208-216.

Dillon, N. (2004). Heterochromatin structure and function. *Biol Cell, 96*(8), 631-637. doi:10.1016/j.biolcel.2004.06.003

Dion, V., & Wilson, J. H. (2009). Instability and chromatin structure of expanded trinucleotide repeats. *Trends Genet, 25*(7), 288-297. doi:10.1016/j.tig.2009.04.007

Disney, M. D., Liu, B., Yang, W. Y., Sellier, C., Tran, T., Charlet-Berguerand, N., & Childs-Disney, J. L. (2012). A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome. *ACS Chem Biol, 7*(10), 1711-1718. doi:10.1021/cb300135h

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dols-Icardo, O., Garcia-Redondo, A., Rojas-Garcia, R., Sanchez-Valle, R., Noguera, A., Gomez-Tortosa, E., . . . Clarimon, J. (2013). Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. *Hum Mol Genet*. doi:10.1093/hmg/ddt460

Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., & Kazazian, H. H., Jr. (1991). Isolation of an active human transposable element. *Science, 254*(5039), 1805-1808.

Dombroski, B. A., Scott, A. F., & Kazazian, H. H., Jr. (1993). Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. *Proc Natl Acad Sci U S A, 90*(14), 6513-6517.

Donnelly, C. J., Zhang, P. W., Pham, J. T., Heusler, A. R., Mistry, N. A., Vidensky, S., . . . Rothstein, J. D. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron, 80*(2), 415-428. doi:10.1016/j.neuron.2013.10.015

Douglas, P. M., & Dillin, A. (2010). Protein homeostasis and aging in neurodegeneration. *J Cell Biol, 190*(5), 719-729. doi:10.1083/jcb.201005144

Dulken, B. W., Leeman, D. S., Boutet, S. C., Hebestreit, K., & Brunet, A. (2017). Single-Cell Transcriptomic Analysis Defines Heterogeneity and Transcriptional Dynamics in the Adult Neural Stem Cell Lineage. *Cell Rep, 18*(3), 777-790. doi:10.1016/j.celrep.2016.12.060

Durr, A. (2010). Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol, 9*(9), 885-894. doi:10.1016/S1474-4422(10)70183-6

Durr, A., Cossee, M., Agid, Y., Campuzano, V., Mignard, C., Penet, C., . . . Koenig, M. (1996). Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N Engl J Med, 335*(16), 1169-1175. doi:10.1056/NEJM199610173351601

Edbauer, D., & Haass, C. (2016). An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. *Curr Opin Neurobiol, 36*, 99-106. doi:10.1016/j.conb.2015.10.009

Eiges, R., Urbach, A., Malcov, M., Frumkin, T., Schwartz, T., Amit, A., . . . Ben-Yosef, D. (2007). Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. *Cell Stem Cell, 1*(5), 568-577. doi:10.1016/j.stem.2007.09.001

Elden, A. C., Kim, H. J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., . . . Gitler, A. D. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature, 466*(7310), 1069-1075. doi:10.1038/nature09320

Erwin, J. A., Marchetto, M. C., & Gage, F. H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. *Nature reviews. Neuroscience, 15*(8), 497-506. doi:10.1038/nrn3730

Evans-Galea, M. V., Carrodus, N., Rowley, S. M., Corben, L. A., Tai, G., Saffery, R., . . . Sarsero, J. P. (2012a). FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Annals of neurology, 71*(4), 487-497. doi:10.1002/ana.22671

Evans-Galea, M. V., Carrodus, N., Rowley, S. M., Corben, L. A., Tai, G., Saffery, R., . . . Sarsero, J. P. (2012b). FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Annals of neurology, 71*(4), 487-497. doi:10.1002/ana.22671

Evrony, G. D., Cai, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., . . . Walsh, C. A. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell, 151*(3), 483-496. doi:10.1016/j.cell.2012.09.035

Evrony, G. D., Lee, E., Park, P. J., & Walsh, C. A. (2016). Resolving rates of mutation in the brain using single-neuron genomics. *Elife, 5*. doi:10.7554/eLife.12966

Farg, M. A., Konopka, A., Ying Soo, K., Ito, D., & Atkin, J. D. (2017). The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in Amyotrophic Lateral Sclerosis. *Hum Mol Genet*. doi:10.1093/hmg/ddx170

Farg, M. A., Sundaramoorthy, V., Sultana, J. M., Yang, S., Atkinson, R. A., Levina, V., . . . Atkin, J. D. (2014). C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet, 23*(13), 3579-3595. doi:10.1093/hmg/ddu068

Farkash, E. A., Kao, G. D., Horman, S. R., & Prak, E. T. (2006). Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic acids research, 34*(4), 1196-1204. doi:10.1093/nar/gkj522

Fend, F., & Raffeld, M. (2000). Laser capture microdissection in pathology. *J Clin Pathol, 53*(9), 666-672.

Fratta, P., Mizielinska, S., Nicoll, A. J., Zloh, M., Fisher, E. M. C., Parkinson, G., & Isaacs, A. M. (2012). C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Scientific reports, 2*. doi:doi:10.1038/srep01016

Fratta, P., Poulter, M., Lashley, T., Rohrer, J. D., Polke, J. M., Beck, J., . . . Mead, S. (2013). Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol, 126*(3), 401-409. doi:10.1007/s00401-013-1147-0

Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K. H., . . . Taylor, J. P. (2015). GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature, 525*(7567), 129-133. doi:10.1038/nature14974

Frost, B., Hemberg, M., Lewis, J., & Feany, M. B. (2014). Tau promotes neurodegeneration through global chromatin relaxation. *Nat Neurosci, 17*(3), 357-366. doi:10.1038/nn.3639

Gendron, T. F., Bieniek, K. F., Zhang, Y. J., Jansen-West, K., Ash, P. E., Caulfield, T., . . . Petrucelli, L. (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol, 126*(6), 829-844. doi:10.1007/s00401-013-1192-8

Gitcho, M. A., Baloh, R. H., Chakraverty, S., Mayo, K., Norton, J. B., Levitch, D., . . . Cairns, N. J. (2008). TDP-43 A315T mutation in familial motor neuron disease. *Annals of neurology, 63*(4), 535-538. doi:10.1002/ana.21344

Goodwin, M., Mohan, A., Batra, R., Lee, K. Y., Charizanis, K., Fernandez Gomez, F. J., . . . Swanson, M. S. (2015). MBNL Sequestration by Toxic RNAs and RNA Misprocessing in the Myotonic Dystrophy Brain. *Cell Rep, 12*(7), 1159-1168. doi:10.1016/j.celrep.2015.07.029

Grabczyk, E., & Usdin, K. (2000). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic acids research, 28*(14), 2815-2822.

Gronskov, K., Brondum-Nielsen, K., Dedic, A., & Hjalgrim, H. (2011). A nonsense mutation in FMR1 causing fragile X syndrome. *European journal of human genetics : EJHG, 19*(4), 489-491. doi:10.1038/ejhg.2010.223

Grun, D., Kester, L., & van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat Methods, 11*(6), 637-640. doi:10.1038/nmeth.2930

Guo, L., & Shorter, J. (2015). It's Raining Liquids: RNA Tunes Viscoelasticity and Dynamics of Membraneless Organelles. *Mol Cell, 60*(2), 189-192. doi:10.1016/j.molcel.2015.10.006

Gustavson, K. H., Blomquist, H. K., & Holmgren, G. (1986). Prevalence of the fragile-X syndrome in mentally retarded boys in a Swedish county. *American journal of medical genetics, 23*(1-2), 581-587.

Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., . . . Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature, 507*(7491), 195-200. doi:10.1038/nature13124

Hagerman, P. J., & Hagerman, R. J. (2004). The fragile-X premutation: a maturing perspective. *Am J Hum Genet, 74*(5), 805-816. doi:10.1086/386296

Hagerman, R. J., Hull, C. E., Safanda, J. F., Carpenter, I., Staley, L. W., O'Connor, R. A., . . . et al. (1994). High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *American journal of medical genetics, 51*(4), 298-308. doi:10.1002/ajmg.1320510404

Halliday, G., Bigio, E. H., Cairns, N. J., Neumann, M., Mackenzie, I. R., & Mann, D. M. (2012). Mechanisms of disease in frontotemporal lobar degeneration: gain of function versus loss of function effects. *Acta Neuropathol, 124*(3), 373-382. doi:10.1007/s00401-012-1030-4

Hancks, D. C., & Kazazian, H. H., Jr. (2016). Roles for retrotransposon insertions in human disease. *Mob DNA, 7*, 9. doi:10.1186/s13100-016-0065-9

Hardy, J., & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science, 297*(5580), 353-356. doi:10.1126/science.1072994

Harms, M., Benitez, B. A., Cairns, N., Cooper, B., Cooper, P., Mayo, K., . . . Cruchaga, C. (2013). C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. *JAMA neurology, 70*(6), 736-741. doi:10.1001/2013.jamaneurol.537

Harms, M. B., Cady, J., Zaidman, C., Cooper, P., Bali, T., Allred, P., . . . Baloh, R. H. (2013). Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. *Neurobiology of aging, 34*(9), 2234 e2213-2239. doi:10.1016/j.neurobiolaging.2013.03.006

Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., . . . Hubbard, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res, 22*(9), 1760-1774. doi:10.1101/gr.135350.111

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., . . . Jones, A. R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature, 489*(7416), 391-399. doi:10.1038/nature11405

Hsiung, G. Y., DeJesus-Hernandez, M., Feldman, H. H., Sengdy, P., Bouchard-Kerr, P., Dwosh, E., . . . Mackenzie, I. R. (2012). Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. *Brain : a journal of neurology, 135*(Pt 3), 709-722. doi:10.1093/brain/awr354

Huang, C., Tong, J., Bi, F., Zhou, H., & Xia, X. G. (2012). Mutant TDP-43 in motor neurons promotes the onset and progression of ALS in rats. *J Clin Invest, 122*(1), 107-118. doi:10.1172/JCI59130

Huber, K. M., Gallagher, S. M., Warren, S. T., & Bear, M. F. (2002). Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proc Natl Acad Sci U S A, 99*(11), 7746-7750. doi:10.1073/pnas.122205699

Huelga, S. C., Vu, A. Q., Arnold, J. D., Liang, T. Y., Liu, P. P., Yan, B. Y., . . . Yeo, G. W. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep, 1*(2), 167-178. doi:10.1016/j.celrep.2012.02.001

Igaz, L. M., Kwong, L. K., Lee, E. B., Chen-Plotkin, A., Swanson, E., Unger, T., . . . Lee, V. M. (2011). Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *J Clin Invest, 121*(2), 726-738. doi:10.1172/JCI44867

Ilieva, H., Polymenidou, M., & Cleveland, D. W. (2009). Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J Cell Biol, 187*(6), 761-772. doi:10.1083/jcb.200908164

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature, 466*(7310), 1129-1133. doi:10.1038/nature09303

Iwahashi, C. K., Yasui, D. H., An, H. J., Greco, C. M., Tassone, F., Nannen, K., . . . Hagerman, P. J. (2006). Protein composition of the intranuclear inclusions of FXTAS. *Brain : a journal of neurology, 129*(Pt 1), 256-271. doi:10.1093/brain/awh650

Jackrel, M. E., DeSantis, M. E., Martinez, B. A., Castellano, L. M., Stewart, R. M., Caldwell, K. A., . . . Shorter, J. (2014). Potentiated Hsp104 variants antagonize diverse proteotoxic misfolding events. *Cell, 156*(1-2), 170-182. doi:10.1016/j.cell.2013.11.047

Jain, A., & Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature, 546*(7657), 243-247. doi:10.1038/nature22386

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science, 293*(5532), 1074-1080. doi:10.1126/science.1063127

Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., & Thornton, C. A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet, 13*(24), 3079-3088. doi:10.1093/hmg/ddh327

Jiang, J., & Cleveland, D. W. (2016). Bidirectional Transcriptional Inhibition as Therapy for ALS/FTD Caused by Repeat Expansion in C9orf72. *Neuron, 92*(6), 1160-1163. doi:10.1016/j.neuron.2016.12.008

Jiang, J., Zhu, Q., Gendron, T. F., Saberi, S., McAlonis-Downes, M., Seelman, A., . . . Lagier-Tourenne, C. (2016). Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. *Neuron, 90*(3), 535-550. doi:10.1016/j.neuron.2016.04.006

Jiang, Y. M., Yamamoto, M., Kobayashi, Y., Yoshihara, T., Liang, Y., Terao, S., . . . Sobue, G. (2005). Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Annals of neurology, 57*(2), 236-251. doi:10.1002/ana.20379

Jin, P., Zarnescu, D. C., Zhang, F., Pearson, C. E., Lucchesi, J. C., Moses, K., & Warren, S. T. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron, 39*(5), 739-747.

Johnson, B. S., McCaffery, J. M., Lindquist, S., & Gitler, A. D. (2008). A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc Natl Acad Sci U S A, 105*(17), 6439-6444. doi:10.1073/pnas.0802082105

Johnson, B. S., Snead, D., Lee, J. J., McCaffery, J. M., Shorter, J., & Gitler, A. D. (2009). TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *The Journal of biological chemistry, 284*(30), 20329-20339. doi:10.1074/jbc.M109.010264

Jovicic, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., . . . Gitler, A. D. (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci, 18*(9), 1226-1229. doi:10.1038/nn.4085

Jubelt, B., & Berger, J. R. (2001). Does viral disease underlie ALS? Lessons from the AIDS pandemic. *Neurology, 57*(6), 945-946.

Kabashi, E., Valdmanis, P. N., Dion, P., Spiegelman, D., McConkey, B. J., Vande Velde, C., . . . Rouleau, G. A. (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature genetics, 40*(5), 572-574. doi:10.1038/ng.132

Kamme, F., Salunga, R., Yu, J., Tran, D. T., Zhu, J., Luo, L., . . . Erlander, M. (2003). Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J Neurosci, 23*(9), 3607-3615.

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., . . . McKnight, S. L. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell, 149*(4), 753-767. doi:10.1016/j.cell.2012.04.017

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science, 303*(5664), 1626-1632. doi:10.1126/science.1089670

Kessler, N. J., Van Baak, T. E., Baker, M. S., Laritsky, E., Coarfa, C., & Waterland, R. A. (2016). CpG methylation differences between neurons and glia are highly conserved from mouse to human. *Hum Mol Genet, 25*(2), 223-232. doi:10.1093/hmg/ddv459

Khan, B. K., Yokoyama, J. S., Takada, L. T., Sha, S. J., Rutherford, N. J., Fong, J. C., . . . Miller, B. L. (2012). Atypical, slowly progressive behavioural variant frontotemporal dementia associated with C9ORF72 hexanucleotide expansion. *Journal of neurology, neurosurgery, and psychiatry, 83*(4), 358-364. doi:10.1136/jnnp-2011-301883

Khosravi, B., Hartmann, H., May, S., Mohl, C., Ederle, H., Michaelsen, M., . . . Edbauer, D. (2017). Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. *Hum Mol Genet, 26*(4), 790-800. doi:10.1093/hmg/ddw432

Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., . . . Taylor, J. P. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature, 495*(7442), 467-473. doi:10.1038/nature11922

Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A., & Marioni, J. C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat Commun, 6*, 8687. doi:10.1038/ncomms9687

King, O. D., Gitler, A. D., & Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res, 1462*, 61-80. doi:10.1016/j.brainres.2012.01.016

Kirkwood, T. B. (2005). Understanding the odd science of aging. *Cell, 120*(4), 437-447. doi:10.1016/j.cell.2005.01.027

Kontrogianni-Konstantopoulos, A., & Bloch, R. J. (2003). The hydrophilic domain of small ankyrin-1 interacts with the two N-terminal immunoglobulin domains of titin. *The Journal of biological chemistry, 278*(6), 3985-3991. doi:10.1074/jbc.M209012200

Kopera, H. C., Larson, P. A., Moldovan, J. B., Richardson, S. R., Liu, Y., & Moran, J. V. (2016). LINE-1 Cultured Cell Retrotransposition Assay. *Methods Mol Biol, 1400*, 139-156. doi:10.1007/978-1-4939-3372-3_10

Koppers, M., Blokhuis, A. M., Westeneng, H. J., Terpstra, M. L., Zundel, C. A., Vieira de Sa, R., . . . Pasterkamp, R. J. (2015). C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Annals of neurology, 78*(3), 426-438. doi:10.1002/ana.24453

Kordasiewicz, H. B., Stanek, L. M., Wancewicz, E. V., Mazur, C., McAlonis, M. M., Pytel, K. A., . . . Cleveland, D. W. (2012). Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron, 74*(6), 1031-1044. doi:10.1016/j.neuron.2012.05.009

Kozlenkov, A., Roussos, P., Timashpolsky, A., Barbu, M., Rudchenko, S., Bibikova, M., . . . Dracheva, S. (2014). Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic acids research, 42*(1), 109-127. doi:10.1093/nar/gkt838

Kramer, N. J., Carlomagno, Y., Zhang, Y. J., Almeida, S., Cook, C. N., Gendron, T. F., . . . Gitler, A. D. (2016). Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. *Science, 353*(6300), 708-712. doi:10.1126/science.aaf7791

Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., . . . Richards, R. I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science, 252*(5013), 1711-1714.

Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nuske, E., Poser, I., . . . Alberti, S. (2015). Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *Elife, 4*, e06807. doi:10.7554/eLife.06807

Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W. W., Morrill, K., . . . Dubnau, J. (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLoS Genet, 13*(3), e1006635. doi:10.1371/journal.pgen.1006635

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., . . . Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research, 44*(W1), W90-97. doi:10.1093/nar/gkw377

Kwiatkowski, T. J., Jr., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., . . . Brown, R. H., Jr. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science, 323*(5918), 1205-1208. doi:10.1126/science.1166066

Ladd, A. C., Brohawn, D. G., Thomas, R. R., Keeney, P. M., Berr, S. S., Khan, S. M., . . . Bennett, J. P. (2017). RNA-seq analyses reveal that cervical spinal cords and anterior motor neurons from amyotrophic lateral sclerosis subjects show reduced expression of mitochondrial DNA-encoded respiratory genes, and rhTFAM may correct this respiratory deficiency. *Brain Res, 1667*, 74-83. doi:10.1016/j.brainres.2017.05.010

Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H. R., . . . Ravits, J. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci U S A, 110*(47), E4530-4539. doi:10.1073/pnas.1318835110

Lagier-Tourenne, C., Polymenidou, M., Hutt, K. R., Vu, A. Q., Baughn, M., Huelga, S. C., . . . Yeo, G. W. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci, 15*(11), 1488-1497. doi:10.1038/nn.3230

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., . . . International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:10.1038/35057062

Langa, K. M., Larson, E. B., Crimmins, E. M., Faul, J. D., Levine, D. A., Kabeto, M. U., & Weir, D. R. (2017). A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012. *JAMA Intern Med, 177*(1), 51-58. doi:10.1001/jamainternmed.2016.6807

Larson, K., Yan, S. J., Tsurumi, A., Liu, J., Zhou, J., Gaur, K., . . . Li, W. X. (2012). Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet, 8*(1), e1002473. doi:10.1371/journal.pgen.1002473

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., . . . Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol, 9*(8), e1003118. doi:10.1371/journal.pcbi.1003118

Lee, E. B., Lee, V. M., & Trojanowski, J. Q. (2012). Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nature reviews. Neuroscience, 13*(1), 38-50. doi:10.1038/nrn3121

Lee, E. B., Porta, S., Michael Baer, G., Xu, Y., Suh, E., Kwong, L. K., . . . Trojanowski, J. Q. (2017). Expansion of the classification of FTLD-TDP: distinct pathology associated with rapidly progressive frontotemporal degeneration. *Acta Neuropathol, 134*(1), 65-78. doi:10.1007/s00401-017-1679-9

Lee, E. B., Russ, J., Jung, H., Elman, L. B., Chahine, L. M., Kremens, D., . . . McCluskey, L. F. (2013). Topography of FUS pathology distinguishes late-onset BIBD from aFTLD-U. *Acta Neuropathol Commun, 1*(9), 1-11. doi:10.1186/2051-5960-1-9

Lee, Y.-B., Chen, H.-J., Peres, João N., Gomez-Deza, J., Attig, J., talekar, M., . . . Shaw, Christopher E. (2013). Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. *Cell reports*.

Lee, Y. B., Chen, H. J., Peres, J. N., Gomez-Deza, J., Attig, J., Stalekar, M., . . . Shaw, C. E. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester

RNA binding proteins, and are neurotoxic. *Cell Rep, 5*(5), 1178-1186. doi:10.1016/j.celrep.2013.10.049

Li, W., Jin, Y., Prazak, L., Hammell, M., & Dubnau, J. (2012). Transposable elements in TDP-43-mediated neurodegenerative disorders. *PLoS One, 7*(9), e44099. doi:10.1371/journal.pone.0044099

Li, W., Lee, M. H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., . . . Nath, A. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med, 7*(307), 307ra153. doi:10.1126/scitranslmed.aac8201

Li, W., Prazak, L., Chatterjee, N., Gruninger, S., Krug, L., Theodorou, D., & Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in Drosophila. *Nat Neurosci, 16*(5), 529-531. doi:10.1038/nn.3368

Li, Y. R., King, O. D., Shorter, J., & Gitler, A. D. (2013). Stress granules as crucibles of ALS pathogenesis. *J Cell Biol, 201*(3), 361-372. doi:10.1083/jcb.201302044

Lin, P. Y., Simon, S. M., Koh, W. K., Folorunso, O., Umbaugh, C. S., & Pierce, A. (2013). Heat shock factor 1 over-expression protects against exposure of hydrophobic residues on mutant SOD1 and early mortality in a mouse model of amyotrophic lateral sclerosis. *Mol Neurodegener, 8*, 43. doi:10.1186/1750-1326-8-43

Lin, Y., Protter, D. S., Rosen, M. K., & Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell, 60*(2), 208-219. doi:10.1016/j.molcel.2015.08.018

Lindquist, S., & Craig, E. A. (1988). The heat-shock proteins. *Annu Rev Genet, 22*, 631-677. doi:10.1146/annurev.ge.22.120188.003215

Ling, J. P., Pletnikova, O., Troncoso, J. C., & Wong, P. C. (2015). TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science, 349*(6248), 650-655. doi:10.1126/science.aab0983

Ling, S. C., Polymenidou, M., & Cleveland, D. W. (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron, 79*(3), 416-438. doi:10.1016/j.neuron.2013.07.033

Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., . . . Ranum, L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science, 293*(5531), 864-867. doi:10.1126/science.1062125

Liu, E. Y., Russ, J., Wu, K., Neal, D., Suh, E., McNally, A. G., . . . Lee, E. B. (2014). C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol, 128*(4), 525-541. doi:10.1007/s00401-014-1286-y

Liu, G. H., Barkho, B. Z., Ruiz, S., Diep, D., Qu, J., Yang, S. L., . . . Izpisua Belmonte, J. C. (2011). Recapitulation of premature ageing with iPSCs from Hutchinson-Gilford progeria syndrome. *Nature, 472*(7342), 221-225. doi:10.1038/nature09879

Liu, X., Li, D., Zhang, W., Guo, M., & Zhan, Q. (2012). Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay. *The EMBO journal, 31*(23), 4415-4427. doi:10.1038/emboj.2012.292

Loesch, D. Z., Huggins, R., Hay, D. A., Gedeon, A. K., Mulley, J. C., & Sutherland, G. R. (1993). Genotype-phenotype relationships in fragile X syndrome: a family study. *Am J Hum Genet, 53*(5), 1064-1073.

Lokireddy, S., Kukushkin, N. V., & Goldberg, A. L. (2015). cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins. *Proc Natl Acad Sci U S A, 112*(52), E7176-7185. doi:10.1073/pnas.1522332112

Lopez Castel, A., Cleary, J. D., & Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol, 11*(3), 165-170. doi:10.1038/nrm2854

Lopez-Gonzalez, R., Lu, Y., Gendron, T. F., Karydas, A., Tran, H., Yang, D., . . . Gao, F. B. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function

and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron, 92*(2), 383-391. doi:10.1016/j.neuron.2016.09.015

Loureiro, J. R., Oliveira, C. L., & Silveira, I. (2016). Unstable repeat expansions in neurodegenerative diseases: nucleocytoplasmic transport emerges on the scene. *Neurobiology of aging, 39*, 174-183. doi:10.1016/j.neurobiolaging.2015.12.007

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Lunnon, K., Smith, R., Hannon, E., De Jager, P. L., Srivastava, G., Volta, M., . . . Mill, J. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci, 17*(9), 1164-1170. doi:10.1038/nn.3782

Luu, L. M., Nguyen, L., Peng, S., Lee, J., Lee, H. Y., Wong, C. H., . . . Zimmerman, S. C. (2016). A Potent Inhibitor of Protein Sequestration by Expanded Triplet (CUG) Repeats that Shows Phenotypic Improvements in a Drosophila Model of Myotonic Dystrophy. *ChemMedChem, 11*(13), 1428-1435. doi:10.1002/cmdc.201600081

MacDonald, M. E., Barnes, G., Srinidhi, J., Duyao, M. P., Ambrose, C. M., Myers, R. H., . . . et al. (1993). Gametic but not somatic instability of CAG repeat length in Huntington's disease. *Journal of medical genetics, 30*(12), 982-986.

MacGowan, D. J., Scelsa, S. N., & Waldron, M. (2001). An ALS-like syndrome with new HIV infection and complete response to antiretroviral therapy. *Neurology, 57*(6), 1094-1097.

Mackenzie, I. R., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., . . . Neumann, M. (2013). Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol, 126*(6), 859-879. doi:10.1007/s00401-013-1181-y

Mackenzie, I. R., Frick, P., Grasser, F. A., Gendron, T. F., Petrucelli, L., Cashman, N. R., . . . Neumann, M. (2015). Quantitative analysis and clinico-pathological correlations of

different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol, 130*(6), 845-861. doi:10.1007/s00401-015-1476-2

Mackenzie, I. R., Frick, P., & Neumann, M. (2014). The neuropathology associated with repeat expansions in the C9ORF72 gene. *Acta Neuropathol, 127*(3), 347-357. doi:10.1007/s00401-013-1232-4

Macleod, D., Charlton, J., Mullins, J., & Bird, A. P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev, 8*(19), 2282-2292.

Mahoney, C. J., Beck, J., Rohrer, J. D., Lashley, T., Mok, K., Shakespeare, T., . . . Warren, J. D. (2012). Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain : a journal of neurology, 135*(Pt 3), 736-750. doi:10.1093/brain/awr361

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., . . . Bates, G. P. (1996). Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell, 87*(3), 493-506.

Mansuroglu, Z., Benhelli-Mokrani, H., Marcato, V., Sultan, A., Violet, M., Chauderlier, A., . . . Bonnefoy, E. (2016). Loss of Tau protein affects the structure, transcription and repair of neuronal pericentromeric heterochromatin. *Sci Rep, 6*, 33047. doi:10.1038/srep33047

Maxwell, P. H., Burhans, W. C., & Curcio, M. J. (2011). Retrotransposition is associated with genome instability during chronological aging. *Proc Natl Acad Sci U S A, 108*(51), 20376-20381. doi:10.1073/pnas.1100271108

May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., . . . Edbauer, D. (2014). C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol, 128*(4), 485-503. doi:10.1007/s00401-014-1329-4

Mayr, C., & Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell, 138*(4), 673-684. doi:10.1016/j.cell.2009.06.016

McConkie-Rosell, A., Lachiewicz, A. M., Spiridigliozzi, G. A., Tarleton, J., Schoenwald, S., Phelan, M. C., . . . Brown, W. T. (1993). Evidence that methylation of the FMR-I locus is responsible for variable phenotypic expression of the fragile X syndrome. *Am J Hum Genet, 53*(4), 800-809.

McCormick, A. L., Brown, R. H., Jr., Cudkowicz, M. E., Al-Chalabi, A., & Garson, J. A. (2008). Quantification of reverse transcriptase in ALS and elimination of a novel retroviral candidate. *Neurology, 70*(4), 278-283. doi:10.1212/01.wnl.0000297552.13219.b4

McMillan, C. T., Russ, J., Wood, E. M., Irwin, D. J., Grossman, M., McCluskey, L., . . . Lee, E. B. (2015). C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. *Neurology, 84*(16), 1622-1630. doi:10.1212/WNL.0000000000001495

Merenstein, S. A., Sobesky, W. E., Taylor, A. K., Riddle, J. E., Tran, H. X., & Hagerman, R. J. (1996). Molecular-clinical correlations in males with an expanded FMR1 mutation. *American journal of medical genetics, 64*(2), 388-394. doi:10.1002/(SICI)1096-8628(19960809)64:2<388::AID-AJMG31>3.0.CO;2-9

Miccoli, L., Biard, D. S., Frouin, I., Harper, F., Maga, G., & Angulo, J. F. (2003). Selective interactions of human kin17 and RPA proteins with chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *Nucleic acids research, 31*(14), 4162-4175.

Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., & Swanson, M. S. (2000a). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *The EMBO journal, 19*(17), 4439-4448. doi:10.1093/emboj/19.17.4439

Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., & Swanson, M. S. (2000b). Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *The EMBO journal, 19*(17), 4439-4448. doi:10.1093/emboj/19.17.4439

Mitchell, J. A., Clay, I., Umlauf, D., Chen, C. Y., Moir, C. A., Eskiw, C. H., . . . Fraser, P. (2012). Nuclear RNA sequencing of the mouse erythroid cell transcriptome. *PLoS One, 7*(11), e49274. doi:10.1371/journal.pone.0049274

Mizielinska, S., Gronke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., . . . Isaacs, A. M. (2014). C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science, 345*(6201), 1192-1194. doi:10.1126/science.1256800

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., . . . Taylor, J. P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell, 163*(1), 123-133. doi:10.1016/j.cell.2015.09.015

Molyneaux, B. J., Arlotta, P., Menezes, J. R., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nature reviews. Neuroscience, 8*(6), 427-437. doi:10.1038/nrn2151

Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., . . . Edbauer, D. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science, 339*(6125), 1335-1338. doi:10.1126/science.1232927

Moseley, M. L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A. K., Daughters, R. S., . . . Ranum, L. P. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nature genetics, 38*(7), 758-769. doi:10.1038/ng1827

Moulignier, A., Moulonguet, A., Pialoux, G., & Rozenbaum, W. (2001). Reversible ALS-like disorder in HIV infection. *Neurology, 57*(6), 995-1001.

Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., & Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature, 435*(7044), 903-910. doi:10.1038/nature03663

Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., & Gage, F. H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature, 468*(7322), 443-446. doi:10.1038/nature09544

Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S., Rees, E., Miyashita, A., . . . St George-Hyslop, P. (2015). ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron, 88*(4), 678-690. doi:10.1016/j.neuron.2015.10.030

Murray, M. E., Bieniek, K. F., Banks Greenberg, M., DeJesus-Hernandez, M., Rutherford, N. J., van Blitterswijk, M., . . . Dickson, D. W. (2013). Progressive amnestic dementia, hippocampal sclerosis, and mutation in C9ORF72. *Acta neuropathologica, 126*(4), 545-554. doi:10.1007/s00401-013-1161-2

Myrick, L., Nakamoto-Kinoshita, M., Lindor, N., Kirmani, S., Cheng, X., & Warren, S. (2014). Fragile X syndrome due to a missense mutation. *European journal of human genetics : EJHG*.

Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., . . . Lee, V. M. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science, 314*(5796), 130-133. doi:10.1126/science.1134108

Nomura, Y., Adachi, N., & Koyama, H. (2007). Human Mus81 and FANCB independently contribute to repair of DNA damage during replication. *Genes Cells, 12*(10), 1111-1122. doi:10.1111/j.1365-2443.2007.01124.x

O'Donnell, W. T., & Warren, S. T. (2002). A decade of molecular studies of fragile X syndrome. *Annual review of neuroscience, 25*, 315-338. doi:10.1146/annurev.neuro.25.112701.142909

O'Rourke, J. G., Bogdanik, L., Muhammad, A. K., Gendron, T. F., Kim, K. J., Austin, A., . . . Baloh, R. H. (2015). C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron, 88*(5), 892-901. doi:10.1016/j.neuron.2015.10.027

O'Rourke, J. G., Bogdanik, L., Yanez, A., Lall, D., Wolf, A. J., Muhammad, A. K., . . . Baloh, R. H. (2016). C9orf72 is required for proper macrophage and microglial function in mice. *Science, 351*(6279), 1324-1329. doi:10.1126/science.aaf1064

Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., & Yuan, J. (2017). Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. *Trends Mol Med, 23*(6), 563-576. doi:10.1016/j.molmed.2017.04.006

Oh, S. Y., He, F., Krans, A., Frazer, M., Taylor, J. P., Paulson, H. L., & Todd, P. K. (2015). RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Hum Mol Genet, 24*(15), 4317-4326. doi:10.1093/hmg/ddv165

Ohki, Y., Wenninger-Weinzierl, A., Hruscha, A., Asakawa, K., Kawakami, K., Haass, C., . . . Schmid, B. (2017). Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration. *Mol Neurodegener, 12*(1), 6. doi:10.1186/s13024-016-0146-8

Pearson, C. E., Nichol Edamura, K., & Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet, 6*(10), 729-742. doi:10.1038/nrg1689

Peters, O. M., Cabrera, G. T., Tran, H., Gendron, T. F., McKeon, J. E., Metterville, J., . . . Brown, R. H., Jr. (2015). Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. *Neuron, 88*(5), 902-909. doi:10.1016/j.neuron.2015.11.018

Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell, 66*(4), 817-822.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K. R., Huelga, S. C., Moran, J., Liang, T. Y., . . . Cleveland, D. W. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci, 14*(4), 459-468. doi:10.1038/nn.2779

Proudfoot, M., Gutowski, N. J., Edbauer, D., Hilton, D. A., Stephens, M., Rankin, J., & Mackenzie, I. R. (2014). Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. *Acta Neuropathol, 127*(3), 451-458. doi:10.1007/s00401-014-1245-7

Prudencio, M., Belzil, V. V., Batra, R., Ross, C. A., Gendron, T. F., Pregent, L. J., . . . Petrucelli, L. (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci, 18*(8), 1175-1182. doi:10.1038/nn.4065

Prudencio, M., Gonzales, P. K., Cook, C. N., Gendron, T. F., Daughrity, L. M., Song, Y., . . . Link, C. D. (2017). Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. *Hum Mol Genet*. doi:10.1093/hmg/ddx233

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., & Trono, D. (2012). The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep, 2*(4), 766-773. doi:10.1016/j.celrep.2012.08.043

Rabin, S. J., Kim, J. M., Baughn, M., Libby, R. T., Kim, Y. J., Fan, Y., . . . Ravits, J. (2010). Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. *Hum Mol Genet, 19*(2), 313-328. doi:10.1093/hmg/ddp498

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc, 8*(11), 2281-2308. doi:10.1038/nprot.2013.143

Reddy, K., Zamiri, B., Stanley, S. Y., Macgregor, R. B., Jr., & Pearson, C. E. (2013). The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni-

and multimolecular RNA G-quadruplex structures. *The Journal of biological chemistry, 288*(14), 9860-9866. doi:10.1074/jbc.C113.452532

Reiss, A. L., Freund, L. S., Baumgardner, T. L., Abrams, M. T., & Denckla, M. B. (1995). Contribution of the FMR1 gene mutation to human intellectual dysfunction. *Nature genetics, 11*(3), 331-334. doi:10.1038/ng1195-331

Renton, A. E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J. R., . . . Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron, 72*(2), 257-268. doi:10.1016/j.neuron.2011.09.010

Richardson, S. R., Morell, S., & Faulkner, G. J. (2014). L1 retrotransposons and somatic mosaicism in the brain. *Annu Rev Genet, 48*, 1-27. doi:10.1146/annurev-genet-120213-092412

Rohrer, J. D., Nicholas, J. M., Cash, D. M., van Swieten, J., Dopper, E., Jiskoot, L., . . . Binetti, G. (2015). Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis. *Lancet Neurol, 14*(3), 253-262. doi:10.1016/S1474-4422(14)70324-2

Rosenbloom, K. R., Sloan, C. A., Malladi, V. S., Dreszer, T. R., Learned, K., Kirkup, V. M., . . . Kent, W. J. (2013). ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic acids research, 41*(Database issue), D56-63. doi:10.1093/nar/gks1172

Rowe, H. M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T. S., Tarabay, Y., . . . Trono, D. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res, 23*(3), 452-461. doi:10.1101/gr.147678.112

Russ, J., Liu, E. Y., Wu, K., Neal, D., Suh, E., Irwin, D. J., . . . Lee, E. B. (2015). Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathol, 129*(1), 39-52. doi:10.1007/s00401-014-1365-0

Saksouk, N., Simboeck, E., & Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin, 8*, 3. doi:10.1186/1756-8935-8-3

Saldi, T. K., Ash, P. E., Wilson, G., Gonzales, P., Garrido-Lecca, A., Roberts, C. M., . . . Link, C. D. (2014). TDP-1, the Caenorhabditis elegans ortholog of TDP-43, limits the accumulation of double-stranded RNA. *The EMBO journal, 33*(24), 2947-2966. doi:10.15252/embj.201488740

Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K., Grant, S., Simpkinson, M., . . . Baloh, R. H. (2013). Targeting RNA Foci in iPSC-Derived Motor Neurons from ALS Patients with a C9ORF72 Repeat Expansion. *Sci Transl Med, 5*(208), 208ra149. doi:10.1126/scitranslmed.3007529

Scekic-Zahirovic, J., Sendscheid, O., El Oussini, H., Jambeau, M., Sun, Y., Mersmann, S., . . . Dupuis, L. (2016). Toxic gain of function from mutant FUS protein is crucial to trigger cell autonomous motor neuron loss. *The EMBO journal, 35*(10), 1077-1097. doi:10.15252/embj.201592559

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J., 3rd. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev, 16*(8), 919-932. doi:10.1101/gad.973302

Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., . . . Charlet-Berguerand, N. (2013). Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep, 3*(3), 869-880. doi:10.1016/j.celrep.2013.02.004

Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R. K., Gattoni, R., . . . Charlet-Berguerand, N. (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *The EMBO journal, 29*(7), 1248-1261. doi:10.1038/emboj.2010.21

Sephton, C. F., Cenik, C., Kucukural, A., Dammer, E. B., Cenik, B., Han, Y., . . . Yu, G. (2011). Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *The Journal of biological chemistry, 286*(2), 1204-1215. doi:10.1074/jbc.M110.190884

Sharma, A., Lyashchenko, A. K., Lu, L., Nasrabady, S. E., Elmaleh, M., Mendelsohn, M., . . . Shneider, N. A. (2016). ALS-associated mutant FUS induces selective motor neuron degeneration through toxic gain of function. *Nat Commun, 7*, 10465. doi:10.1038/ncomms10465

Sharma, K., Schmitt, S., Bergner, C. G., Tyanova, S., Kannaiyan, N., Manrique-Hoyos, N., . . . Simons, M. (2015). Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci, 18*(12), 1819-1831. doi:10.1038/nn.4160

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., . . . Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A, 111*(51), E5593-5601. doi:10.1073/pnas.1419161111

Shiga, A., Ishihara, T., Miyashita, A., Kuwabara, M., Kato, T., Watanabe, N., . . . Onodera, O. (2012). Alteration of POLDIP3 splicing associated with loss of function of TDP-43 in tissues affected with ALS. *PLoS One, 7*(8), e43120. doi:10.1371/journal.pone.0043120

Simon-Sanchez, J., Dopper, E. G., Cohn-Hokke, P. E., Hukema, R. K., Nicolaou, N., Seelaar, H., . . . van Swieten, J. C. (2012). The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. *Brain : a journal of neurology, 135*(Pt 3), 723-735. doi:10.1093/brain/awr353

Smeets, H. J., Smits, A. P., Verheij, C. E., Theelen, J. P., Willemsen, R., van de Burgt, I., . . . Oostra, B. A. (1995). Normal phenotype in two brothers with a full FMR1 mutation. *Hum Mol Genet, 4*(11), 2103-2108.

Smith, J., Calidas, D., Schmidt, H., Lu, T., Rasoloson, D., & Seydoux, G. (2016). Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. *Elife, 5*. doi:10.7554/eLife.21337

Smith, K. T., & Workman, J. L. (2012). Chromatin proteins: key responders to stress. *PLoS Biol, 10*(7), e1001371. doi:10.1371/journal.pbio.1001371

Snowden, J. S., Rollinson, S., Thompson, J. C., Harris, J. M., Stopford, C. L., Richardson, A. M., . . . Pickering-Brown, S. M. (2012). Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. *Brain : a journal of neurology, 135*(Pt 3), 693-708. doi:10.1093/brain/awr355

Sofola, O. A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., . . . Botas, J. (2007). RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron, 55*(4), 565-571. doi:10.1016/j.neuron.2007.07.021

Sreedharan, J., Blair, I. P., Tripathi, V. B., Hu, X., Vance, C., Rogelj, B., . . . Shaw, C. E. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science, 319*(5870), 1668-1672. doi:10.1126/science.1154584

Srinivasan, K., Friedman, B. A., Larson, J. L., Lauffer, B. E., Goldstein, L. D., Appling, L. L., . . . Hansen, D. V. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nat Commun, 7*, 11295. doi:10.1038/ncomms11295

Steele, A. J., Al-Chalabi, A., Ferrante, K., Cudkowicz, M. E., Brown, R. H., Jr., & Garson, J. A. (2005). Detection of serum reverse transcriptase activity in patients with ALS and unaffected blood relatives. *Neurology, 64*(3), 454-458. doi:10.1212/01.WNL.0000150899.76130.71

Stepto, A., Gallo, J. M., Shaw, C. E., & Hirth, F. (2014). Modelling C9ORF72 hexanucleotide repeat expansion in amyotrophic lateral sclerosis and frontotemporal dementia. *Acta Neuropathol, 127*(3), 377-389. doi:10.1007/s00401-013-1235-1

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., . . . Cedar, H. (2009). Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol, 16*(5), 564-571. doi:10.1038/nsmb.1594

Strong, M. J., Abrahams, S., Goldstein, L. H., Woolley, S., McLaughlin, P., Snowden, J., . . . Turner, M. R. (2017). Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotroph Lateral Scler Frontotemporal Degener, 18*(3-4), 153-174. doi:10.1080/21678421.2016.1267768

Strong, M. J., Volkening, K., Hammond, R., Yang, W., Strong, W., Leystra-Lantz, C., & Shoesmith, C. (2007). TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol Cell Neurosci, 35*(2), 320-327. doi:10.1016/j.mcn.2007.03.007

Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W. Y., . . . Disney, M. D. (2014). Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron, 83*(5), 1043-1050. doi:10.1016/j.neuron.2014.07.041

Sun, Z., Diaz, Z., Fang, X., Hart, M. P., Chesi, A., Shorter, J., & Gitler, A. D. (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol, 9*(4), e1000614. doi:10.1371/journal.pbio.1000614

Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., & Warren, S. T. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet, 1*(6), 397-400.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., . . . Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science, 324*(5929), 930-935. doi:10.1126/science.1170116

Tan, H., Qurashi, A., Poidevin, M., Nelson, D. L., Li, H., & Jin, P. (2012). Retrotransposon activation contributes to fragile X premutation rCGG-mediated neurodegeneration. *Hum Mol Genet, 21*(1), 57-65. doi:10.1093/hmg/ddr437

Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., & Singer, R. H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol, 128*(6), 995-1002.

Tao, Z., Wang, H., Xia, Q., Li, K., Li, K., Jiang, X., . . . Ying, Z. (2015). Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet, 24*(9), 2426-2441. doi:10.1093/hmg/ddv005

Team, R. C. (2015). R: A Language and Environment for Statistical Computing.

Toledo, J. B., Van Deerlin, V. M., Lee, E. B., Suh, E., Baek, Y., Robinson, J. L., . . . Trojanowski, J. Q. (2013). A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. *Alzheimer's & dementia : the journal of the Alzheimer's Association*. doi:10.1016/j.jalz.2013.06.003

Toledo, J. B., Van Deerlin, V. M., Lee, E. B., Suh, E., Baek, Y., Robinson, J. L., . . . Trojanowski, J. Q. (2014). A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. *Alzheimers Dement, 10*(4), 477-484 e471. doi:10.1016/j.jalz.2013.06.003

Tollervey, J. R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., . . . Ule, J. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci, 14*(4), 452-458. doi:10.1038/nn.2778

Tran, H., Almeida, S., Moore, J., Gendron, T. F., Chalasani, U., Lu, Y., . . . Gao, F. B. (2015). Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron, 87*(6), 1207-1214. doi:10.1016/j.neuron.2015.09.015

Tsai, K. J., Yang, C. H., Fang, Y. H., Cho, K. H., Chien, W. L., Wang, W. T., . . . Shen, C. K. (2010). Elevated expression of TDP-43 in the forebrain of mice is sufficient to cause neurological and pathological phenotypes mimicking FTLD-U. *J Exp Med, 207*(8), 1661-1673. doi:10.1084/jem.20092164

Tsurumi, A., & Li, W. X. (2012). Global heterochromatin loss: a unifying theory of aging? *Epigenetics, 7*(7), 680-688. doi:10.4161/epi.20540

Upton, K. R., Gerhardt, D. J., Jesuadian, J. S., Richardson, S. R., Sanchez-Luque, F. J., Bodea, G. O., . . . Faulkner, G. J. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. *Cell, 161*(2), 228-239. doi:10.1016/j.cell.2015.03.026

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M. E., Heckman, M. G., Diehl, N. N., . . . Rademakers, R. (2013). Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *Lancet Neurol, 12*(10), 978-988. doi:10.1016/S1474-4422(13)70210-2

Van Deerlin, V. M., Leverenz, J. B., Bekris, L. M., Bird, T. D., Yuan, W., Elman, L. B., . . . Yu, C. E. (2008). TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol, 7*(5), 409-416. doi:10.1016/S1474-4422(08)70071-1

Van Den Bosch, L., Van Damme, P., Bogaert, E., & Robberecht, W. (2006). The role of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. *Biochim Biophys Acta, 1762*(11-12), 1068-1082. doi:10.1016/j.bbadis.2006.05.002

Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., & Gorbunova, V. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun, 5*, 5011. doi:10.1038/ncomms6011

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K. J., Nishimura, A. L., Sreedharan, J., . . . Shaw, C. E. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science, 323*(5918), 1208-1211. doi:10.1126/science.1165942

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., & Mandel, J. L. (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. *Nature, 349*(6310), 624-626. doi:10.1038/349624a0

Voigt, A., Herholz, D., Fiesel, F. C., Kaur, K., Muller, D., Karsten, P., . . . Schulz, J. B. (2010). TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS One, 5*(8), e12247. doi:10.1371/journal.pone.0012247

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., . . . Zoldos, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic acids research, 44*(12), 5615-5628. doi:10.1093/nar/gkw159

Waite, A. J., Baumer, D., East, S., Neal, J., Morris, H. R., Ansorge, O., & Blake, D. J. (2014a). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiology of aging*. doi:10.1016/j.neurobiolaging.2014.01.016

Waite, A. J., Baumer, D., East, S., Neal, J., Morris, H. R., Ansorge, O., & Blake, D. J. (2014b). Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiology of aging, 35*(7), 1779 e1775-1779 e1713. doi:10.1016/j.neurobiolaging.2014.01.016

Wang, B., Matsuoka, S., Ballif, B. A., Zhang, D., Smogorzewska, A., Gygi, S. P., & Elledge, S. J. (2007). Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science, 316*(5828), 1194-1198. doi:10.1126/science.1139476

Wang, I. F., Tsai, K. J., & Shen, C. K. (2013). Autophagy activation ameliorates neuronal pathogenesis of FTLD-U mice: a new light for treatment of TARDBP/TDP-43 proteinopathies. *Autophagy, 9*(2), 239-240. doi:10.4161/auto.22526

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic acids research, 41*(Web Server issue), W77-83. doi:10.1093/nar/gkt439

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., . . . Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature, 454*(7200), 126-130. doi:10.1038/nature06992

Watts, G. D., Wymer, J., Kovach, M. J., Mehta, S. G., Mumm, S., Darvish, D., . . . Kimonis, V. E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature genetics, 36*(4), 377-381. doi:10.1038/ng1332

Webb, T., Thake, A., & Todd, J. (1986). Twelve families with fragile X(q27). *Journal of medical genetics, 23*(5), 400-406.

Webster, C. P., Smith, E. F., Shaw, P. J., & De Vos, K. J. (2017). Protein Homeostasis in Amyotrophic Lateral Sclerosis: Therapeutic Opportunities? *Front Mol Neurosci, 10*, 123. doi:10.3389/fnmol.2017.00123

Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., . . . Trotti, D. (2014). Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron, 84*(6), 1213-1225. doi:10.1016/j.neuron.2014.12.010

Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, A. R., Nakamori, M., Cheng, S. H., . . . Thornton, C. A. (2012). Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature, 488*(7409), 111-115. doi:10.1038/nature11362

Wils, H., Kleinberger, G., Janssens, J., Pereson, S., Joris, G., Cuijt, I., . . . Kumar-Singh, S. (2010). TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proc Natl Acad Sci U S A, 107*(8), 3858-3863. doi:10.1073/pnas.0912417107

Wood, J. G., Jones, B. C., Jiang, N., Chang, C., Hosier, S., Wickremesinghe, P., . . . Helfand, S. L. (2016). Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. *Proc Natl Acad Sci U S A, 113*(40), 11277-11282. doi:10.1073/pnas.1604621113

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., . . . Rogaeva, E. (2013). Hypermethylation of the CpG Island Near the G4C2 Repeat in ALS with a C9orf72 Expansion. *Am J Hum Genet, 92*(6), 981-989. doi:10.1016/j.ajhg.2013.04.017

Xiang, S., Kato, M., Wu, L. C., Lin, Y., Ding, M., Zhang, Y., . . . McKnight, S. L. (2015). The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei. *Cell, 163*(4), 829-839. doi:10.1016/j.cell.2015.10.040

Xu, Y. F., Gendron, T. F., Zhang, Y. J., Lin, W. L., D'Alton, S., Sheng, H., . . . Petrucelli, L. (2010). Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *J Neurosci, 30*(32), 10851-10859. doi:10.1523/JNEUROSCI.1630-10.2010

Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D. L., . . . Jin, P. (2013). Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci U S A, 110*(19), 7778-7783. doi:10.1073/pnas.1219643110

Yamanaka, K., Chun, S. J., Boillee, S., Fujimori-Tonou, N., Yamashita, H., Gutmann, D. H., . . . Cleveland, D. W. (2008). Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nat Neurosci, 11*(3), 251-253. doi:10.1038/nn2047

Yang, L., Embree, L. J., Tsai, S., & Hickstein, D. D. (1998). Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *The Journal of biological chemistry, 273*(43), 27761-27764.

Yang, X., Han, H., De Carvalho, D. D., Lay, F. D., Jones, P. A., & Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell, 26*(4), 577-590. doi:10.1016/j.ccr.2014.07.028

Young, L. C., McDonald, D. W., & Hendzel, M. J. (2013). Kdm4b histone demethylase is a DNA damage response protein and confers a survival advantage following gamma-irradiation. *The Journal of biological chemistry, 288*(29), 21376-21388. doi:10.1074/jbc.M113.491514

Zeisel, A., Munoz-Manchado, A. B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., . . . Linnarsson, S. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science, 347*(6226), 1138-1142. doi:10.1126/science.aaa1934

Zhang, D., Iyer, L. M., He, F., & Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. *Front Genet, 3*, 283. doi:10.3389/fgene.2012.00283

Zhang, W., Li, J., Suzuki, K., Qu, J., Wang, P., Zhou, J., . . . Belmonte, J. C. (2015). Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science, 348*(6239), 1160-1163. doi:10.1126/science.aaa1356

Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keeffe, S., . . . Wu, J. Q. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci, 34*(36), 11929-11947. doi:10.1523/JNEUROSCI.1860-14.2014

Zhang, Y., Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., . . . Barres, B. A. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron, 89*(1), 37-53. doi:10.1016/j.neuron.2015.11.013

Zhao, W., Beers, D. R., Henkel, J. S., Zhang, W., Urushitani, M., Julien, J. P., & Appel, S. H. (2010). Extracellular mutant SOD1 induces microglial-mediated motoneuron injury. *Glia, 58*(2), 231-243. doi:10.1002/glia.20919

Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., . . . Ranum, L. P. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A, 108*(1), 260-265. doi:10.1073/pnas.1013343108

Zu, T., Liu, Y., Banez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., . . . Ranum, L. P. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A, 110*(51), E4968-4977. doi:10.1073/pnas.1315438110